

Korzekwa, Dominika (2016) *Filling in gaps of *Drosophila melanogaster* urate degradation metabolic pathway using metabolomics approaches: towards the core metabolome of the fruit fly*. PhD thesis.

<http://theses.gla.ac.uk/7706/>

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given



University  
of Glasgow

**Filling in gaps of *Drosophila melanogaster* urate  
degradation metabolic pathway using metabolomics  
approaches: towards the core metabolome of the  
fruit fly**

**Dominika Korzekwa**

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

**Institute of Molecular, Cell and Systems Biology  
College of Medical, Veterinary and Life Sciences  
University of Glasgow  
Glasgow G12 8QQ**

March 2016



# Abstract

The primary goal of systems biology is to integrate complex omics data, and data obtained from traditional experimental studies in order to provide a holistic understanding of organismal function. One way of achieving this aim is to generate genome-scale metabolic models (GEMs), which contain information on all metabolites, enzyme-coding genes, and biochemical reactions in a biological system. *Drosophila melanogaster* GEM has not been reconstructed to date.

Constraint-free genome-wide metabolic model of the fruit fly has been reconstructed in our lab, identifying gaps, where no enzyme was identified and metabolites were either only produced or consumed. The main focus of the work presented in this thesis was to develop a pipeline for efficient gap filling using metabolomics approaches combined with standard reverse genetics methods, using 5-hydroxyisourate hydrolase (5-HIUH) as an example. 5-HIUH plays a role in urate degradation pathway. Inability to degrade urate can lead to inborn errors of metabolism (IEMs) in humans, including hyperuricemia.

Based on sequence analysis *Drosophila* CG30016 gene was hypothesised to encode 5-HIUH. CG30016 knockout flies were examined to identify Malpighian tubules phenotype, and shortened lifespan might reflect kidney disorders in hyperuricemia in humans. Moreover, LC-MS analysis of mutant tubules revealed that CG30016 is involved in purine metabolism, and specifically urate degradation pathway. However, the exact role of the gene has not been identified, and the complete method for gap filling has not been developed. Nevertheless, thanks to the work presented here, we are a step closer towards the development of a gap-filling pipeline in *Drosophila melanogaster* GEM. Importantly, the areas that require further optimisation were identified and are the focus of future research. Moreover, LC-MS analysis confirmed that tubules rather than the whole fly were more suitable for metabolomics analysis of purine metabolism.

Previously, Dow/Davies lab has generated the most complete tissue-specific transcriptomic atlas for *Drosophila* – FlyAtlas.org, which provides data on gene expression across multiple tissues of adult fly and larva. FlyAtlas revealed that transcripts of many genes are enriched in specific *Drosophila* tissues, and that it is possible to deduce the functions of individual tissues within the fly. Based on FlyAtlas data, it has become clear that the fly (like other metazoan species) must be considered as a set of tissues, each

with its own distinct transcriptional and functional profile. Moreover, it revealed that for about 30% of the genome, reverse genetic methods (i.e. mutation in an unknown gene followed by observation of phenotype) are only useful if specific tissues are investigated. Based on the FlyAtlas findings, we aimed to build a primary tissue-specific metabolome of the fruit fly, in order to establish whether different *Drosophila* tissues have different metabolomes and if they correspond to tissue-specific transcriptome of the fruit fly (FlyAtlas.org). Different fly tissues have been dissected and their metabolome elucidated using LC-MS. The results confirmed that tissue metabolomes differ significantly from each other and from the whole fly, and that some of these differences can be correlated to the tissue function. The results illustrate the need to study individual tissues as well as the whole organism. It is clear that some metabolites that play an important role in a given tissue might not be detected in the whole fly sample because their abundance is much lower in comparison to other metabolites present in all tissues, which prevent the detection of the tissue-specific compound.



Credit: Lisa Hanawalt ([http://www.nytimes.com/2012/01/15/opinion/sunday/fruit-flies-and-love.html?\\_r=0](http://www.nytimes.com/2012/01/15/opinion/sunday/fruit-flies-and-love.html?_r=0))

## ***Fruit flies***

by Nicholas Friedman

*In crooked paths, they waft  
through August, ping-pong from  
fruit to fruit, gang aft*

*agley, then rest on plum  
or Brandywine to lay  
their careful schemes: in sum,*

*thousands of eggs per day.  
They curse each coffee cup  
to drink the cream, and pay*

*with life for that one taste.  
But I, being a man,  
have countless tries to waste*

*in winging rot to rot,  
pursuing finer things.  
The hands of others swat*

*me, rising up and up—  
a driven little fly  
bound for a coffee cup.*

# Acknowledgements

My deepest gratitude goes to my sister, Paulina, and my wonderful, loving and accepting friends, without whom I would have not been able to finish this work. My sister has always been by my side, be it silently, actively supporting or in conflict. At the end, thanks to her, I have been growing up stronger yet more sensitive every day, and I am who I am today. She was there when I needed her most to support me, and enabled me to finish my PhD.

I thank my supervisors Prof. Julian Dow and Prof. Shireen Davies, without whom this work would not be possible. I am particularly grateful for their patience and openness. These taught me not only to persevere in my research, but also in life in a broad sense.

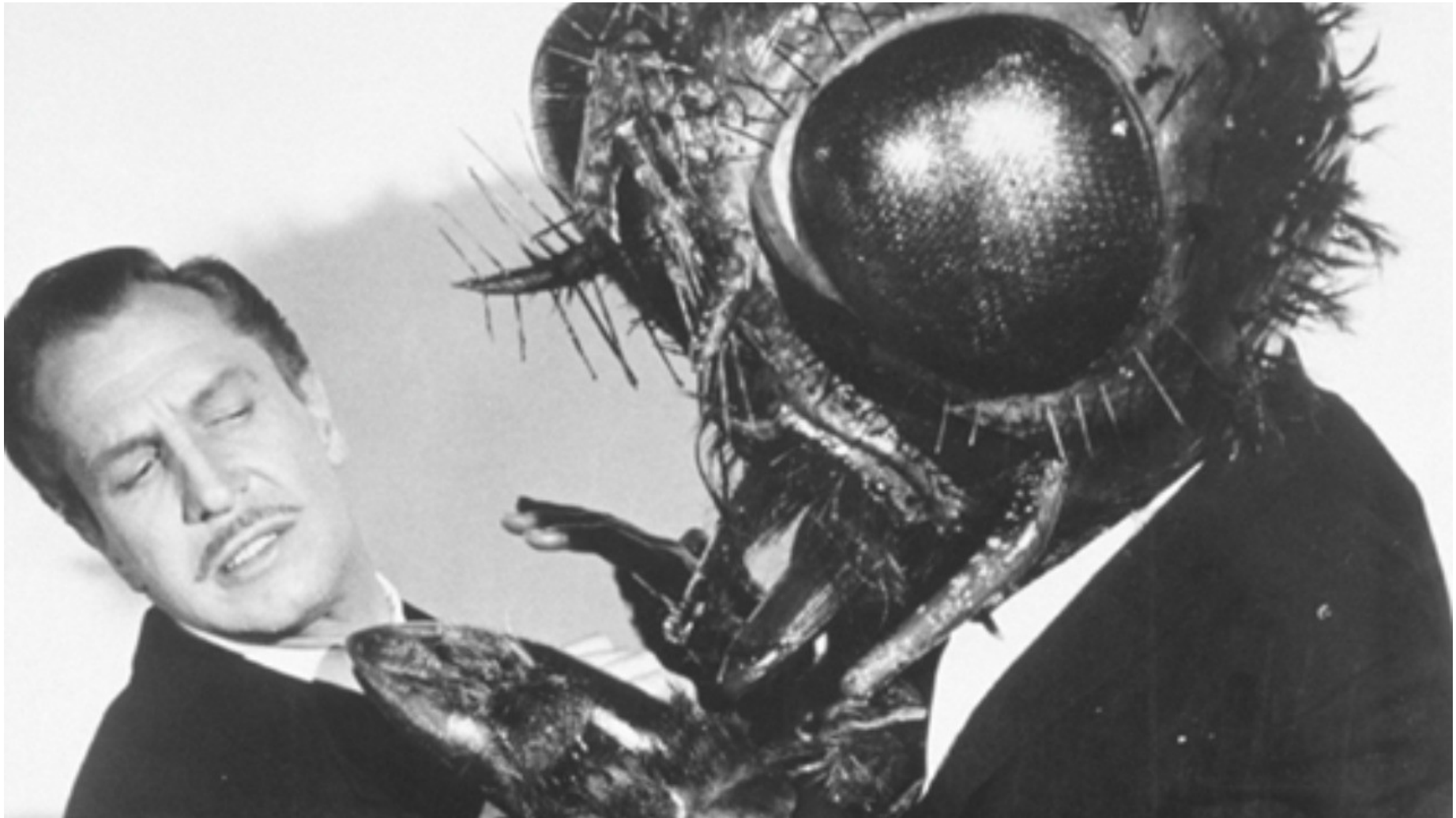
I would like to thank our collaborators Dr David Watson and Dr Venkat Chintapalli for sharing their data and expertise on tissue-specific metabolomes of the fruit fly.

This project was hugely supported by all Dow/Davies lab members, and especially Pablo, Bhoomi and Selim. I thank everyone for being there, sharing knowledge and challenging me every day – both in my research – and on a personal level. This has been an amazing lesson. Bhoomi, thank you for all these moments of crying, laughing, dancing, singing, and finally, living together in our last months of the PhD.

Among my friends, I would like to mention Klaudia, Kalina, Rosalind, Rosh, Francesco, Karolina, Romain, John, Paul, Pol, Sue, and Pernilla, who helped me survive this time through constantly supporting me and making my life outside the PhD an incredible, exciting and challenging journey. Special thanks go to Francesco, who supported and motivated me in the last month of my write-up process, by sitting across the desk from me in libraries, kitchens, bedrooms and cafés, providing food and wine, planning and going on hiking trips, and accepting my mood swings.

I would like to thank my parents for always pushing me to achieve set goals, and for having taught me what hard work meant. As much as I struggled with these, they provided the major force allowing me to complete my PhD.

And finally, to all the fruit flies sacrificed during the process of my research. I shall never hurt another one of you!



From 'Le Retour de la Mouche', a film by Edward L. Bernds (1959). 'Whatever you give away today or think or say or do, will multiply about tenfold and then return to you.'

## **Author's Declaration**

This thesis is the result of my own work and includes nothing, except results presented in chapter 5.2.1, which is the outcome of work done in collaboration. The results obtained in collaboration with Dr David Watson (University of Strathclyde) and Dr Venkat Chintapalli (University of Glasgow) are clearly highlighted in the text and are presented for the purpose of comparison with my own work.

# Table of Contents

<b>ABSTRACT .....</b>	<b>2</b>
<b>ACKNOWLEDGEMENTS.....</b>	<b>5</b>
<b>AUTHOR'S DECLARATION .....</b>	<b>7</b>
<b>TABLE OF CONTENTS .....</b>	<b>8</b>
<b>LIST OF TABLES AND FIGURES .....</b>	<b>16</b>
<b>LIST OF ABBREVIATIONS .....</b>	<b>21</b>
<b>1 INTRODUCTION .....</b>	<b>25</b>
<b>1.1 <i>Drosophila</i> as a genetic model.....</b>	<b>25</b>
1.1.1 Drosophila genome.....	25
1.1.2 <i>Drosophila</i> molecular-genetic tools .....	26
1.1.3 Transposable elements.....	27
1.1.4 GAL4-UAS system.....	28
<b>1.2 <i>Drosophila</i> as a model organism.....</b>	<b>29</b>
1.2.1 Genetics and epigenetics .....	30
1.2.2 Developmental biology.....	30
1.2.3 Biomedicine .....	31
1.2.3.1 IEMs and <i>Drosophila</i> .....	33
1.2.3.2 The need to study individual tissues.....	34
<b>1.3 Metabolomics .....</b>	<b>35</b>
1.3.1 What is metabolomics used for? .....	36
1.3.1.1 Medicine .....	36
1.3.1.2 Microbial research .....	37
1.3.1.3 Food science .....	37

1.3.1.4	Animal research .....	37
1.3.2	Metabolomics technologies.....	37
1.3.2.1	NMR .....	38
1.3.2.2	MS-based platforms.....	39
1.3.2.3	LC-MS for metabolomics.....	41
1.3.3	Sample preparation for LC-MS analysis .....	42
1.3.4	LC-MS data processing.....	44
1.3.4.1	Data pre-treatment.....	45
1.3.4.2	Peak annotation and matching.....	45
1.3.4.3	Metabolite identification .....	46
1.3.4.4	Statistical analysis of LC-MS data.....	47
1.3.4.5	Biological interpretation of LC-MS data.....	48
1.3.4.6	Tools for metabolomics data processing .....	48
1.3.5	Limitations and challenges of LC-MS for metabolomics .....	49
1.3.5.1	Metabolite identification .....	49
1.3.5.2	Data interpretation .....	51
1.3.5.3	Data visualisation.....	52
1.3.5.4	Integration of metabolomics data .....	52
<b>1.4</b>	<b>Systems biology .....</b>	<b>53</b>
1.4.1	Functional genomics for systems biology .....	54
1.4.2	Metabolomics for functional genomics .....	55
1.4.3	Metabolomics – a systems biology tool.....	56
1.4.4	Genome-scale metabolic map reconstruction.....	57
1.4.4.1	Gaps in GEMs.....	58
1.4.4.2	Filling in the gaps .....	59
1.4.5	Towards global metabolic model of <i>Drosophila melanogaster</i> .....	60
<b>1.5</b>	<b>Project aims .....</b>	<b>61</b>
<b>2</b>	<b>MATERIALS AND METHODS.....</b>	<b>63</b>
<b>2.1</b>	<b>Drosophila maintenance and stocks .....</b>	<b>63</b>
2.1.1	Fly stocks.....	63
2.1.2	Fly maintenance.....	63



2.1.3	Fly mating .....	65
<b>2.2</b>	<b><i>Drosophila</i> tissue dissection .....</b>	<b>65</b>
<b>2.3</b>	<b>RNA extraction .....</b>	<b>66</b>
2.3.1	RNA extraction using whole flies and whole larvae .....	66
2.3.2	RNA extraction using Malpighian tubules .....	67
2.3.3	Complementary DNA synthesis .....	67
2.3.4	Quantification of nucleic acid .....	67
<b>2.4</b>	<b>Oligonucleotide synthesis .....</b>	<b>68</b>
<b>2.5</b>	<b>Polymerase chain reaction (PCR) .....</b>	<b>68</b>
2.5.1	Standard PCR .....	68
2.5.2	Quantitative reverse-transcriptase PCR (qPCR) .....	69
2.5.3	Data analysis .....	70
<b>2.6</b>	<b>Agarose gel electrophoresis .....</b>	<b>71</b>
<b>2.7</b>	<b>PCR/Gel purification .....</b>	<b>71</b>
<b>2.8</b>	<b>Molecular cloning .....</b>	<b>71</b>
2.8.1	Bacterial strains and plasmids .....	72
2.8.2	Construct generation .....	72
2.8.3	Cell-free protein expression system .....	72
2.8.3.1	Cloning primers and PCR amplification .....	73
2.8.4	TOPO pET100 expression system .....	73
2.8.4.1	Cloning primers and PCR amplification .....	74
2.8.5	DES TOPO cloning .....	74
2.8.6	Cloning into pUAST plasmid .....	75
2.8.6.1	Cloning primers and PCR amplification .....	76
2.8.7	Transformation of <i>E.coli</i> cells .....	76
2.8.8	Purification of plasmid .....	77
2.8.9	Validation of cloning products .....	77
2.8.9.1	PCR .....	77
2.8.9.2	Restriction digest .....	78
2.8.9.3	Sequencing .....	78

2.8.10	Pilot expression of <i>Drosophila</i> protein in BL21 cells .....	78
<b>2.9</b>	<b><i>Drosophila</i> S2 Cell Techniques .....</b>	<b>79</b>
2.9.1	Maintenance.....	79
2.9.2	Transient transfection.....	79
<b>2.10</b>	<b>Protein purification and analysis .....</b>	<b>80</b>
2.10.1	Protein purification.....	80
2.10.2	Changing protein buffer and concentration .....	80
2.10.3	Bradford assay .....	80
2.10.4	SDS-PAGE separation .....	81
2.10.5	Western blotting.....	81
<b>2.11</b>	<b>Immunocytochemistry .....</b>	<b>82</b>
2.11.1	ICC of S2 cells.....	82
2.11.2	ICC of <i>Drosophila</i> Malpighian tubules.....	83
<b>2.12</b>	<b>Imaging .....</b>	<b>83</b>
2.12.1	Fluorescent imaging.....	83
2.12.2	Polarised light microscopy .....	83
<b>2.13</b>	<b>Stress and survival experiments .....</b>	<b>85</b>
2.13.1	Oxidative stress survival assay .....	85
2.13.2	Longevity assay .....	85
<b>2.14</b>	<b>Fluid secretion assay .....</b>	<b>86</b>
2.14.1	Principles of the assay.....	86
2.14.2	Collection of fluid for LC-MS .....	87
<b>2.15</b>	<b>Haemolymph extraction .....</b>	<b>87</b>
<b>2.16</b>	<b>Metabolomics analysis .....</b>	<b>87</b>
2.16.1	Metabolite extraction .....	87
2.16.2	Liquid chromatography-Mass spectrometry (LC-MS) .....	88
2.16.3	Data processing .....	88
<b>2.17</b>	<b>CD-spectroscopy.....</b>	<b>90</b>

2.18	Bioinformatics tools for <i>CG30016</i> and <i>urate oxidase</i> sequence analysis.....	90
<b>3</b>	<b>URATE DEGRADATION PATHWAY.....</b>	<b>92</b>
<b>3.1</b>	<b>Introduction .....</b>	<b>92</b>
3.1.1	Evolution of urate degradation pathway.....	93
3.1.2	<i>Drosophila</i> enzymes involved in urate degradation pathway.....	94
3.1.2.1	<i>Drosophila</i> Urate oxidase.....	95
3.1.2.2	5-Hydroxyisourate hydrolase.....	96
<b>3.2</b>	<b>Results .....</b>	<b>97</b>
3.2.1	<i>Drosophila melanogaster</i> <i>CG30016</i> gene .....	97
3.2.2	<i>Drosophila</i> <i>CG30016</i> and <i>Urate oxidase</i> mutants.....	101
3.2.3	<i>CG30016</i> and <i>Urate oxidase</i> expression at different developmental stages in <i>Drosophila</i> larvae .....	103
3.2.4	Expression of <i>CG30016</i> and <i>Urate oxidase</i> in mutant L3 larvae .....	104
<b>3.3</b>	<b>Discussion .....</b>	<b>105</b>
<b>4</b>	<b>PHENOTYPIC MANIFESTATIONS IN <i>DROSOPHILA MELANOGASTER</i> MUTANTS OF URATE DEGRADATION PATHWAY .....</b>	<b>107</b>
<b>4.1</b>	<b>Introduction .....</b>	<b>107</b>
4.1.1	<i>Drosophila melanogaster</i> Malpighian tubules.....	107
4.1.1.1	Waste and fluid secretion .....	108
4.1.1.2	Tubule phenotypes .....	109
4.1.1.3	Tubules in biomedicine .....	109
4.1.2	Physiological role of urate .....	110
4.1.2.1	Benefits of uric acid .....	111
4.1.2.2	Urate as an inducer of oxidative stress.....	111
<b>4.2</b>	<b>Results .....</b>	<b>112</b>
4.2.1	Urate crystals .....	112
4.2.2	Inflated ureter phenotype .....	114
4.2.3	Inflated abdomen phenotype .....	116
4.2.4	Tubule secretion rate.....	116

4.2.5	Oxidative stress survival .....	117
4.2.5.1	Oxidative stress induced by 1% H <sub>2</sub> O <sub>2</sub> .....	117
4.2.5.2	Oxidative stress induced by 10 mM paraquat .....	120
4.2.6	Longevity assay .....	122
<b>4.3</b>	<b>Discussion .....</b>	<b>124</b>
4.3.1	Urate Crystals.....	124
4.3.2	Inflated tubule and abdomen phenotype and fluid secretion rate .....	125
4.3.3	Is urate an antioxidant or a pro-oxidative factor in the fruit fly? .....	126
<b>5</b>	<b>METABOLOMICS ANALYSIS IN <i>DROSOPHILA MELANOGASTER</i> .....</b>	<b>128</b>
<b>5.1</b>	<b>Introduction .....</b>	<b>128</b>
5.1.1	How do single-gene mutations affect the fly metabolome? .....	130
<b>5.2</b>	<b>Results .....</b>	<b>131</b>
5.2.1	Tissue-specific metabolomes of <i>Drosophila melanogaster</i> .....	131
5.2.1.1	Tissue metabolomes differ from each other and from the whole fly .....	132
5.2.1.2	Most abundant metabolites – amino acids .....	136
5.2.1.3	Metabolomes of Malpighian tubules.....	140
5.2.1.4	Metabolomes of the gut .....	142
5.2.1.5	Metabolome of <i>Drosophila melanogaster</i> head and central nervous system 145	
5.2.2	Metabolomics analysis of mutants of purine metabolism .....	148
5.2.2.1	Adult fly metabolomes.....	148
5.2.2.2	Adult fly metabolomes – purine metabolism .....	153
5.2.2.3	Mutant larva metabolomes .....	158
5.2.3	Metabolomes of secreted fluid.....	163
<b>5.3</b>	<b>Discussion .....</b>	<b>166</b>
<b>6</b>	<b>IN VITRO EXPRESSION OF <i>DROSOPHILA MELANOGASTER</i> CG30016 AND URATE OXIDASE .....</b>	<b>170</b>
<b>6.1</b>	<b>Introduction .....</b>	<b>170</b>
6.1.1	Choosing the expression system.....	170

6.1.1.1	Cell-free expression system .....	171
6.1.1.2	Expression using <i>E. coli</i> .....	171
6.1.1.3	<i>Drosophila</i> S2 cells .....	172
6.1.2	Detection of urate degradation intermediates.....	172
6.1.2.1	Detection using LC-MS.....	172
6.1.2.2	Detection using CD-spectroscopy .....	173
<b>6.2</b>	<b>Results .....</b>	<b>174</b>
6.2.1	Cell-free expression system .....	174
6.2.1.1	Construct generation .....	174
6.2.1.2	Expression and protein detection.....	175
6.2.1.3	Enzyme assay and Liquid chromatography – mass spectrometry (LC-MS).....	176
6.2.2	Expression of <i>UO</i> and <i>CG30016</i> in S2 cells.....	183
6.2.2.1	Construct generation .....	183
6.2.2.2	Transient transfection and its efficiency.....	183
6.2.2.3	Detection of expressed protein .....	185
6.2.2.4	LC-MS of S2 cells .....	185
6.2.3	Expression of <i>UO</i> and <i>CG30016</i> in <i>E. coli</i> .....	186
6.2.3.1	Construct generation: PCR, primers, gel purification, ligation.....	186
6.2.3.2	Expression and protein detection.....	187
6.2.3.3	Enzyme assay and CD - spectroscopy .....	189
<b>6.3</b>	<b>Discussion .....</b>	<b>192</b>
6.3.1	Problems with protein expression .....	192
6.3.2	Metabolite detection: LC-MS versus CD-spectroscopy.....	193
6.3.2.1	LC-MS detection.....	193
6.3.2.2	CD-spectroscopy detection.....	193
6.3.3	Future work .....	194
<b>7</b>	<b>CONCLUSIONS AND FUTURE WORK.....</b>	<b>195</b>
7.1	Development of a method for gap filling of <i>Drosophila metabolic</i> global metabolic model .....	195
7.2	<i>CG30016</i> is involved in urate degradation pathway .....	196

7.2.1	QPCR analysis.....	196
7.2.2	Sequence analysis and literature review .....	196
7.2.3	Phenotypic manifestations .....	197
7.2.4	Metabolomics analysis.....	197
7.2.5	<i>In vitro</i> experiments.....	199
7.2.6	Future work .....	199
<b>7.3</b>	<b>Different <i>Drosophila</i> tissues have different metabolic profiles .....</b>	<b>200</b>
7.3.1	LC-MS measurement of secreted fluid .....	201
7.3.2	Future work .....	201
<b>7.4</b>	<b>Final conclusion.....</b>	<b>202</b>
<b>8</b>	<b>REFERENCES .....</b>	<b>204</b>
	<b>APPENDICES .....</b>	<b>227</b>

# List of Tables and Figures

Fig. 1-1 Induction of transgenes using GAL4-UAS binary expression system, from [36].....	29
Table 1-1 Comparison between MS-based and NMR metabolomics including advantages and disadvantages of both methods. ....	39
Fig. 1-2 Scheme of metabolomics workflow. Adapted from [136]. ....	44
Table 2-1 Fly lines used in this study. Each fly line used is described including its ID, genotype, brief description and source. ....	64
Table 2-2 Developmental <i>Drosophila</i> larval stages. ....	65
Table 2-3 Standard PCR thermal cycling conditions.....	69
Table 2-4 QPCR reaction components. ....	70
Table 2-5 Bacterial strains and vectors used in the study.....	72
Fig 2.1 Map of pF25A ICE T7 Flexi® Vector.....	73
Fig 2.2 Map of pET101/D-TOPO® plasmid.....	74
Fig 2.3 Map of pMT/V5-His-TOPO® vector.....	75
Fig 2.4 Map of pUAST vector.....	76
Table 2-6 Reaction components and the quantities used for transient transfection of S2 cells. ....	79
Table 2-7 Sample components and quantity for SDS-PAGE. ....	81
Fig.2-5 Polarised light microscope configuration. ....	84
Table 2-8 Types of food preparations used in oxidative stress survival assay. ....	85
Fig. 2-6 Fluid secretion assay performed in <i>Drosophila</i> [352].....	86
Fig. 2-7 IDEOM: data processing pipeline used for LC-MS data processing, adapted from [158]. .....	89
Fig 3-1 Formation and breakdown of urate. urate to S-allantoin takes several seconds.....	92
Fig. 3-2 Phylogenetic genome comparison of urate degradation enzymes. ....	94

Fig. 3-3 Tissues-specific expression of Urate oxidase in <i>Drosophila melanogaster</i> . ....	95
Table 3-1 Characteristic TRP features. ....	96
Fig. 3-4 Tissue-specific expression of <i>CG30016</i> in <i>Drosophila melanogaster</i> . ....	98
Fig. 3-5 Sequence alignment of <i>CG30016</i> with mouse (A) and human (B) transthyretin genes. .	99
Fig. 3-6 Similarity metrics of <i>CG30016</i> gene with 5-HIU hydrolases from other species, revealing 30-41% sequence identity. ....	100
Fig. 3-7 Sequence alignment of <i>CG30016</i> with soybean 5-hydroxyisourate hydrolase. ....	101
Fig. 3-8 Prediction of intracellular localisation of <i>CG30016</i> . ....	101
Fig. 3-9 Expression of <i>hbs</i> gene in BDSC 17767 line (control flies). ....	102
Fig. 3-10 QPCR validation of mutant flies BDSC 18554 and BDSC 18814. ....	102
Fig. 3-11 Expression of <i>UO</i> and <i>CG30016</i> genes throughout larval development. ....	104
Fig. 3-12 Expression of <i>UO</i> and <i>CG30016</i> genes in L3 mutant larvae. ....	105
Fig 4-1 <i>Drosophila melanogaster</i> morphology. ....	108
Table 4-1 Human kidney diseases modelled in <i>Drosophila</i> and fly genes homologous to human genes associated with the disease. ....	110
Fig. 4-2 Malpighian tubules of <i>Drosophila</i> Cantos S L1-L3 larvae and urate crystals. ....	113
Fig. 4-3 Urate crystals in <i>Drosophila</i> BDSC 17767 L1-L3 larval Malpighian tubules. ....	113
Fig 4-4 Urate crystals in <i>Drosophila</i> Malpighian tubules. ....	114
Fig. 4-5 Inflated ureter phenotype in BDSC 18554 mutant flies. ....	115
Fig. 4-6 Inflated abdomen phenotype (A) and distorted abdominal pattern (B) in BDSC 18554 mutant flies. ....	116
Fig. 4-7 Basal secretion rate (nl/min) comparison between BDSC 18554 and BDAC 17767 flies. .....	117
Fig. 4-8 Oxidative stress survival assay induced by 1% H <sub>2</sub> O <sub>2</sub> . ....	119



Fig. 4-9 Survival assay in oxidative stress conditions induced by 10 mM paraquat.....	121
Fig. 4-10 Longevity assay, comparison between mutant flies and controls.....	123
Fig. 5-1 The principle of single-gene knockout metabolomics.....	130
Fig. 5-2 Hierarchical clustering analysis (A) and PCA (B) of tissue metabolomes. ....	134
Fig. 5-3 Results of one-way ANOVA. ....	135
Fig. 5-4 The top 25 most abundant metabolites by area response in the whole fly and ten fly tissues. . ....	136
Table 5-1 The most enriched metabolites across all tissues. ....	138
Fig. 5-5 The most abundant metabolites by area response in Malpighian tubules. ....	140
Table 5-2 Metabolites enriched in Malpighian tubules (the highest tissue enrichment value for MTs). ....	141
Fig. 5-6 The most abundant metabolites in fly midgut. ....	142
Table 5-3 Metabolites enriched in the gut (the highest tissue enrichment value in the gut)....	143
Fig. 5-7 The most abundant metabolites in <i>Drosophila melanogaster</i> head. ....	145
Table 5-4 Metabolites enriched in central nervous system (the highest tissue enrichment value in CNS). ....	146
Table 5-5 The most abundant metabolites in whole <i>CG30016</i> mutant flies.....	149
Table 5-6 The most abundant metabolites in Malpighian tubules of adult <i>CG30016</i> mutant flies. .....	150
Fig. 5-8 Metabolic pathways affected by the knockout of <i>CG30016</i> in whole flies. ....	151
Fig. 5-9 Results of the PCA analysis. ....	152
Fig. 5-10 Changes in purine metabolism in the whole fly caused by the knockout of <i>CG30016</i> . .....	153
Fig. 5-11 Changes in purine metabolism in adult Malpighian tubules caused by the knockout of <i>CG30016</i> . ....	155

Fig. 5-11 Changes in purine metabolism in adult Malpighian tubules caused by the knockout of <i>CG30016</i> . .....	156
Fig. 5-12 KEGG map of <i>Drosophila melanogaster</i> purine metabolism showing changes in purine metabolism in L3 larvae caused by the knockout of <i>CG30016</i> . .....	158
Fig. 5-13 Changes in purine metabolism in whole larvae caused by the knockout of <i>UO</i> and <i>CG30016</i> . .....	159
Fig. 5-14 Changes in purine metabolism in larval posterior Malpighian tubules caused by the knockout of <i>UO</i> and <i>CG30016</i> . .....	160
Fig. 5-15 Changes in purine metabolism in larval anterior Malpighian tubules caused by the knockout of <i>UO</i> and <i>CG30016</i> . .....	161
Table 5-7 The most abundant metabolites in secreted fluid. ....	163
Fig. 5-16 Metabolites involved in purine metabolism enriched in secreted fluid in comparison to Schneider's medium. ....	165
Fig. 6-1 Urate degradation pathway, enzyme assay. ....	169
Table 6-1 Comparison of three expression systems used in the study including their advantages and disadvantages. ....	170
Fig. 6-2 LC-MS mass peaks of the urate degradation intermediates: urate, 5-HIU, OHCU and allantoin, from [384]. ....	172
Fig. 6-3 Urate degradation pathway monitored using CD-spectroscopy. ....	173
Fig. 6-4 Transgenic construct validation by PCR. ....	174
Fig. 6-5 Detection of protein expressed using the cell-free expression system. ....	175
Figure 6-6 Ion chromatograms and fragmentation spectra of urate and allantoin standards. .	177
Table 6-2 Compounds detected in LC-MS: name of each compound, formula, mass as well as type of detection are provided. ....	178
Table 6-3 LC-MS results of enzyme assay. ....	179
Fig. 6-7 Validation of transgenic constructs by restriction digest. ....	182

Fig. 6-8 Transient transfection of S2 cells. ....	184
Fig. 6-9 Immunocytochemistry assay of S2 cells transfected with constructs containing <i>UO</i> and <i>CG30016</i> genes. ....	184
Table 6-4 LC-MS results of S2 cells transfected with three construct.....	185
Fig. 6-10 Validation of transgenic constructs by restriction digest. ....	186
Fig. 6-11 Determination of optimal expression time and temperature using Western blotting. .....	187
Fig. 5-12 Circular dichroism time course measurement of 0.4 U of UO conversion of 0.1 mM Urate.....	188
Fig. 6-13 Circular dichroism of 0.02 U of UO conversion of 0.1 mM urate. ....	189
Fig. 6-14 CD Ellipticity values at different wavelengths following addition of 0.02 U of UO to 0.1 mM urate.....	190

## List of Abbreviations

<b>2D</b>	two dimensional
<b>5-HIU</b>	5-hydroxyisourate
<b>5-HIUH</b>	5-hydroxyisourate hydrolase
<b>ANOVA</b>	analysis of variance
<b>ATP</b>	adenosine triphosphate
<b>BCAA</b>	branch chain amino acid
<b>BDSC</b>	Bloomington Drosophila Stock Center
<b>BLAST</b>	basic local alignment search tool
<b>bp</b>	base pairs
<b>BSA</b>	bovine serum albumin
<b>C</b>	degrees Celsius
<b>CD</b>	circular dichroism
<b>cDNA</b>	complementary DNA
<b>CE</b>	capillary electrophoresis
<b>CFM</b>	competitive fragmentation modelling
<b>CIP</b>	calf intestine phosphatase
<b>CNS</b>	central nervous system
<b>CS</b>	Canton S
<b>CSM</b>	complete Schneider's medium
<b>Da</b>	Daltons
<b>DAPI</b>	4',6-diamidino-2-phenylindole
<b>DEPC</b>	diethylpyrocabonate
<b>DNA</b>	deoxyribonucleic acid
<b>dNTP</b>	deoxyribonucleotide triphosphates
<b>DR</b>	dietary restriction
<b>dUTP</b>	deoxyuridine triphosphate
<b>EDTA</b>	ethylenediaminetetraacetic acid
<b>EI</b>	electron ionisation
<b>EIC</b>	extracted ion chromatogram

<b>ESI</b>	electrospray ionization
<b>FCS</b>	fetal calf serum
<b>FDA</b>	Food and Drug Administration
<b>FIA</b>	flow injection analysis
<b>FTI</b>	fourier transform infra-red
<b>FXS</b>	fragile X syndrome
<b>GC</b>	gas spectrometry
<b>GEM</b>	genome-scale metabolic model
<b>GFP</b>	green fluorescent protein
<b>GS</b>	goat serum
<b>HBS</b>	HEPES buffered saline
<b>HCA</b>	hierarchical cluster analysis
<b>HEPES</b>	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
<b>HILIC</b>	hydrophilic interaction liquid chromatography
<b>HMDB</b>	Human Metabolome Database
<b>HPLC</b>	high performance LC
<b>ICC</b>	immunocytochemistry
<b>ICD</b>	isotope coded derivatisation
<b>IEM</b>	inborn error of metabolism
<b>IPTG</b>	isopropyl $\beta$ -D-thiogalactoside
<b>Kb</b>	kilo base pairs = 1000 bp
<b>KEGG</b>	Kyoto Encyclopaedia of Genes and Genomes
<b>L-DOPA</b>	L-3,4-dihydroxyphenylalanine
<b>L1</b>	first instar larva
<b>L2</b>	second instar larva
<b>L3</b>	third instar larva
<b>L3E</b>	early third instar larva
<b>LB</b>	lysogeny broth
<b>LC</b>	liquid chromatography
<b>Mb</b>	mega base pairs = 1,000,000 bp
<b>mRNA</b>	messenger RNA

<b>MS</b>	mass spectrometry
<b>MS/MS</b>	tandem MS
<b>MTs</b>	Malpighian tubules
<b>NMR</b>	nuclear magnetic resonance
<b>OHCU</b>	2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline
<b>PAGE</b>	polyacrylamide gel electrophoresis
<b>pBLAST</b>	protein BLAST
<b>PBS</b>	phosphate buffered saline
<b>PBT</b>	PBS + Triton X-100
<b>PCA</b>	principal component analysis
<b>PCR</b>	polymerase chain reaction
<b>PF</b>	Glasgow Universtiy Polyomics Facility
<b>Pfu</b>	Pyrococcus furiosus
<b>PLS-DA</b>	partial-least-squares discriminant analysis
<b>PPM</b>	mass error
<b>PTS2</b>	peroxisome targeting signal 2
<b>QC</b>	quality control
<b>QEMS</b>	Q-Exactive MS
<b>qPCR</b>	quantitative PCR
<b>QSRR</b>	quantitative structure retention relationship
<b>R-allantoin</b>	racemic allantoin
<b>RNA</b>	ribonucleic acid
<b>RNAi</b>	RNA interference
<b>Rnase</b>	ribonuclease
<b>RP</b>	reverse phase
<b>ROS</b>	reactive oxygen species
<b>RP</b>	reverse phase
<b>RSD</b>	relative standard deviation
<b>RT-PCR</b>	real-time PCR
<b>RTm</b>	room temperature
<b>S-Allantoin</b>	dextrorotary allantoin

<b>S2</b>	Drosophila Schneider 2 cells
<b>SDS</b>	sodium dodecyl sulfate
<b>SEM</b>	standard error of the mean
<b>SFC</b>	supercritical fluid chromatography
<b>SU</b>	Strathclyde University
<b>Taq</b>	Thermus aquaticus
<b>TOF</b>	time-of-flight
<b>TRP</b>	transthyretin-related protein
<b>UAS</b>	upstream activating sequence
<b>UDG</b>	uracil-DNA glycosylase
<b>UHPLC</b>	ultra high performance LC
<b>UO</b>	urate oxidase
<b>UP</b>	ultra pure
<b>UV</b>	ultra violet
<b>V-ATPase</b>	vacuolar-type-ATPases
<b>WF</b>	whole fly

# 1 Introduction

## 1.1 *Drosophila* as a genetic model

*Drosophila melanogaster* (fruit fly) was first introduced to the scientific research in 1901 by William Castle at Harvard University and subsequently used as an experimental organism by T.H. Morgan in 1906, who described for the first time the *white* mutant. Ever since, researchers from different fields, including genetics, developmental, cancer research and many more, have extensively used *Drosophila* in their investigations.

*Drosophila* is currently one of the most popular genetic model organisms used in biomedical research. There are many reasons for its success. First of all, its short life cycle (around 2 weeks) means that it grows and reproduces very fast in comparison to other model organisms, such as mouse. Importantly, the maintenance of fly lines is inexpensive (\$20 per year per fly line) and relatively easy. Moreover, *Drosophila* research has an ethical advantage (3Rs: 'replacement, reduction, refinement') [1] over mammalian genetic models.

Another advantage is the availability of *Drosophila* genome sequence published in 2000 [2], as well as a number of stock centers that can provide strains that carry a mutation in a given gene. As a result, it only takes several months from obtaining mutant flies to starting lab work. Finally, availability of fruit fly specific online resources, such as FlyBase (<http://flybase.org/>) [3], FlyAtlas [4], Homophila [5] and many more, is a great advantage for *Drosophila* research. FlyBase provides information about fly gene sequences, RNAi (RNA interference) knockdown lines, mutant alleles, human disease homologues, etc. It is also linked to all *Drosophila* websites, including stock centers. It has recently introduced a new tool for searching fly models of human diseases (Human Disease Model Reports), which is an initiative to encourage collaborations with scientists from different fields, who do not have the expertise in fly research [3]. FlyAtlas.org provides a tissue-specific transcriptome of *Drosophila melanogaster*, which revealed how important looking at individual tissues is, as opposed to the whole organism alone. Based on FlyAtlas data, tissue-specific activities and functions in the fly have been identified [6].



### 1.1.1 *Drosophila* genome

*Drosophila* genome, containing four chromosomes, is much smaller than human genome, which makes genetic experiments simpler. The size of the genome is 180 Mb (mega base pairs), comprising around 130 Mb of euchromatin and around 16,000 genes, which is comparable to the number of genes in the human genome. Out of these 13,000 encode proteins [2]. Moreover, around 70% of *Drosophila* genes have human homologs [7], and over 65% of known human disease genes have fly orthologues [5, 8]. However, even though the fly genome is sequenced and well studied, nearly a quarter of all genes have no known function. Moreover, only around 20% of genes were named and researched in depth [9]. This results in the so-called phenotype gap, where there are no causative connections between the genome and the phenotype variations [10].

One of the advantages of *Drosophila* genetics is the presence of balancer chromosomes. They prevent recombination and enable the maintenance of deleterious mutations in stable fly populations. This is due to the presence of chromosomal inversions in the balancers, which prevent the formation of crossovers. As a result, there is no recombination between homologous chromosomes, and the mutation is maintained. Moreover, they carry visible dominant marker genes (such as curly wings or stubbly bristles), which enable selection of progeny, which inherited the balancer. Balancer chromosomes also carry at least one recessive lethal locus in order to maintain mutations in the stocks.

### 1.1.2 *Drosophila* molecular-genetic tools

There are multiple genetic tools available in the field of *Drosophila* research, including P-element transgenesis, Gal4-UAS system and RNAi. These can be used to manipulate the expression of target genes in the fly at different life stages and in different tissues.

Obtaining a mutant for a given fly gene could not be easier due to the presence of public-domain stock centres. Around half of all fly genes have mutants available in stock centres and were generated by systematic mutagenesis programs[11]. Mutant fly lines include classical mutants, generated by X-ray or chemical mutagenesis or spontaneously arising. Moreover, RNAi lines are available for nearly all genes, and P-element insertional mutants targeting many fly genes have been generated and maintained in the stock centres [12]. Driver lines are also available, allowing expression of transgenic constructs in specific

tissues/cells at a given time.

### 1.1.3 Transposable elements

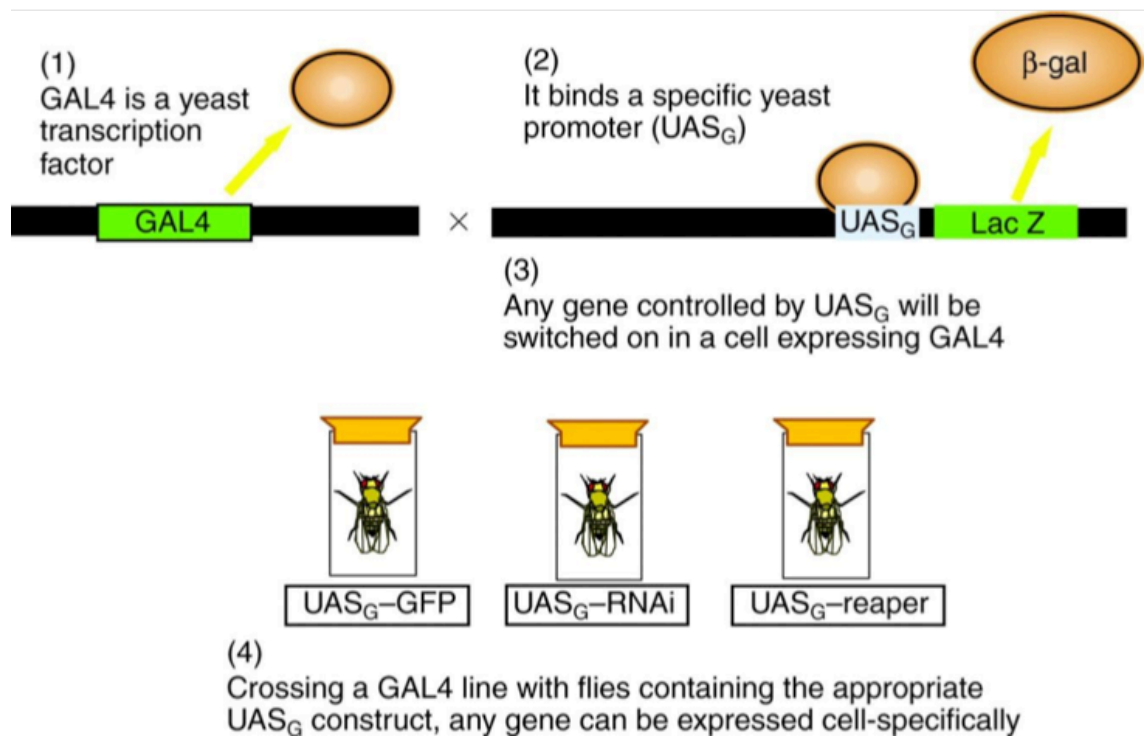
The most common spontaneous mutations, that alter gene expression in the fruit fly, are triggered by the insertion of transposable elements into the genome [13-16]. Several classes of transposable elements naturally occur in *Drosophila*. They represent around 10% of the fly genome [17]. P-elements belong to the most widely studied transposable elements in the fly. They can be used to insert genetic material into the germ-line cells at high frequency [12]. This system provides an efficient mechanism for the insertion of DNA fragments into the germ-line, without rearrangements and providing stable inheritance in future fly generations. This is possible due to certain features of the P-elements. They are DNA fragments, which are 2.9 kilobases (kb) and contain a short 31 base pair (bp) terminal sequence (terminal inverted repeats). These repeats flank four open reading frames, which encode transposase [18]. Expression of the transposase is required for the mobilization and insertion of the P-element into the recipient genome. Transposase can only be expressed in the germ-line because it is dependent on the presence of a specific splicing event that takes place in germ-line cells [19]. Moreover, P-element transposition can be regulated by the repressor protein, which is present in some germ-line cell types [20]. P-elements are particularly useful mutagens as they can carry other DNA fragments including RNAi constructs and visible markers, such as *miniwhite* gene. Genes can be inserted into a specific site of the P-element in order not to affect the expression of transposase. However, most commonly, defective P-elements are generated that do not express transposase and are stable. These are co-injected with the complete P-elements (expressing transposase) into *Drosophila* embryos, resulting in their insertion into the fly genome. P-elements tend to insert on a 5' end of genes and so the expression of genes they carry occurs in a similar pattern to the nearest gene. This is referred to as an enhancer trap [21]. The exact insertion site depends on the size and orientation of the P-element [22]. Because the insertion sites are effectively tagged with known sequences and the full sequence of *Drosophila* is now available, it is possible to identify the insertion point by plasmid rescue or inverse PCR (Polymerase Chain Reaction) [23]. Finally, because of the large size of the P-element, the insertion itself (regardless of the inserted sequence) can be mutagenic.

Gene Distribution Project was started by the *Berkereley Drosophila Genome Project* in order to obtain P-element insertions in all fly genes [24, 25]. However, it is predicted that around 30% of all P-elements are inserted within 400bp of transcription start-site[24, 25] and cluster in hot-spots (areas of the genome that 'attract' multiple P-elements). In order to obtain insertions in as many genes as possible, other transposable elements have been used, which show less insertional specificity [26], including *piggyBacs* [27-29] and Minos element [30, 31].

#### **1.1.4 GAL4-UAS system**

P-element mutagenesis system is commonly used in conjunction with GAL4-UAS binary expression system [32] (Figure 1-1). The system employs GAL4 driver, which is the yeast transcriptional activator, and a UAS (Upstream Activation Sequence) line, which carries a gene of interest containing GAL4 binding sites. GAL4 can activate gene expression in *Drosophila*, where the GAL4 binding sites are present in the promoter region of the gene [33]. Moreover, the GAL4 binding site has been modified in order to allow binding of GAL4 with high affinity [34]. For the UAS-gene to be expressed, the line has to be crossed to GAL4 driver line. Otherwise, the gene remains silent due to the absence of the activator. This way, the phenotypic manifestations of the gene expression can be observed in the progeny. Importantly, in case where the expression of the UAS-cargo is lethal, it is possible to maintain the transgenic lines, and observe the lethal phenotype in the progeny.

Moreover, a library of transgenic lines expressing GAL4 activator in different fly tissues and cells has been generated [32], and at present there are multiple activator-expressing lines available from stock centers. This allows for the time- and tissue-specific regulation of the gene expression. The system is particularly useful for experimental study due to the availability of vectors allowing generation of UAS-RNAi lines, GAL4 lines, or UAS lines with GFP (green fluorescent protein) or epitope tags. This allows generation of new constructs targeting any gene of interest and expressed in any cell/tissue type. The use of UAS-RNAi lines is particularly useful for reverse genetics studies. Moreover, UAS lines with epitope tags can be used to generate overexpressor lines. For a review on RNAi techniques see [35].



**Fig. 1-1 Induction of transgenes using GAL4-UAS binary expression system, from [36].** (1) One part of the system employs a fly line carrying GAL4 (a yeast transcriptional activator) in their genome. (2) Another fly line has UAS (Upstream Activation Sequence) in its genome, which contains GAL4 binding sites. (3) UAS is fused to a gene of interests, whose expression is controlled by UAS, and gets switched on upon GAL4 binding to the UAS. (4) The gene of interest can be expressed in a time- and tissue- specific manner in the progeny produced by crossing these two fly lines.

## 1.2 *Drosophila* as a model organism

As previously mentioned, *Drosophila* is one of the most popular model organisms. This is due to its short life cycle, inexpensive and easy maintenance, complete genome sequence, availability of genome-scale reagent collections, and numerous optimized transgenic tools and molecular techniques for gene manipulation (reviewed in [37]) [3, 38, 39]. Among publicly-available available *Drosophila* resources (reviewed in [40]) are complementary DNA (cDNA) resource [41], insertional mutagenesis collections [24, 42], RNA interference (RNAi) collections for targeted gene knockdown in cells [43, 44] or in flies [11]).

*Drosophila melanogaster* is a powerful model for studying human physiology, development and disease. Despite obvious morphological differences between the fly and the human, there is a functional analogy between many internal organ systems of *Drosophila* and vertebrates, including humans. Moreover, molecular processes that

control and conduct cellular and physiological functions are conserved between both species [38]. Finally, it has been estimated that more than 65% of human disease genes have functional homologues in the fruit fly [5, 8]. Moreover, Chintapalli *et al.* showed that many of these homologues are expressed or enriched in fly tissues that have the same role of the equivalent human tissue [6]. Since evolutionarily conserved genes usually share molecular functions, *Drosophila* has been used as a model for studying molecular mechanisms underlying genetic disturbance, development and behaviour [38]. As a result *Drosophila melanogaster* has greatly contributed to our understanding of mammalian biology [45]. Moreover, it has been widely used as a model in geneticists, developmental biology, and biomedicine [39].

### **1.2.1 Genetics and epigenetics**

*Drosophila* has been used as a model for studying genetics for decades. It contributed to the understanding of the chromosomal basis of inheritance [46, 47], and facilitated the studies on X-rays mutagenesis [48]. Moreover, studies in the fly revealed that unequal crossovers led to duplications and deletions [49] and are the cause of many human genetic disorders [50]. Previous studies have shown that epigenetic gene regulation and chromatin modification are a result of genes that influence position-effect variegation [51], which have been discovered thanks to fruit fly research [52].

### **1.2.2 Developmental biology**

Studies in *Drosophila* have provided a lot of insight into organismal development. These include research on Notch signalling system [53], which mediates cell-cell interactions, and when defective, can lead to cancer and other human diseases [8, 54]. Another example is investigation of homeotic genes in the fly [55]. Homeobox genes influence body plan in flies, and have many functions in all higher eukaryotes [56]. Their aberrations lead to cancer in humans [57]. Other developmental signalling pathways were studied in *Drosophila*, including Toll, Wnt, Hedgehog, and BMP/TGF $\beta$  pathways [58]. Moreover, fruit fly research contributed the understanding of the human nervous system development and wiring [59], as well as heart development [60]. These studies greatly aided the understanding of vertebrate development (including human development) and related diseases as well as cancer [61].

### 1.2.3 Biomedicine

*Drosophila melanogaster* is the most commonly used model organism for studying human pathology. It is a particularly attractive model organism in biomedicine because it allows time- and cost-effective research, and is not limited by ethical considerations, as is the case of other model organisms, such as mouse.

Interestingly, the fly reproduces phenotypic manifestations of certain human diseases, such as renal diseases [62, 63] or neurological disorders [64]. Manipulations of *Drosophila* gene homologous of human disease genes have been shown to result in phenotypic manifestations in different tissues, including MTs (Malpighian tubules) [65-68], abdomen [69], antennae and bristle [70]. Moreover, phenotypes related to physiological functions have been reported, such as altered fluid secretion rate [71-74], and altered longevity and survival in different stress conditions [75, 76]. This makes it particularly suitable for studying human conditions (more on *Drosophila* phenotypes in Chapter 4).

Due to the features mentioned above, *Drosophila melanogaster* is the most commonly used model organism for studying human pathology. It contributed to the understanding of many neurological disorders [45], including Huntington disease [77], Parkinson disease [64], neurodegenerative diseases [78], amyloid lateral sclerosis [79], and narcolepsy and restless leg syndrome [80]. It has also been used for studying cardiovascular disease [81] and cardiac function [82], obesity [83], and drug addiction [84]. Moreover, fruit fly research aided our understanding of animal metabolism and metabolic disorders [85, 86]. Finally, *Drosophila* is a particularly powerful tool for studying vector-borne diseases. These diseases present a huge threat to human health, and mosquito is one of the most dangerous vector species. Mosquito is a vector for West Nile virus, yellow fever, dengue fever, and malaria [87]. Majority of fruit fly research can be translated to other insect species, and aid our understanding of insect genetics and molecular biology. Hence, *Drosophila* has a great impact on development of new methods for controlling mosquito populations. For example, many insecticides have used as a defence against disease vectors. Their mechanism has been studied in *Drosophila* for years, which aided their specificity and effectiveness in disease vectors, including mosquitos [88, 89]. Recently, metabolomics has been used to elucidate how a common insecticide, permethrin, affects insect metabolome and survival. It revealed that lipid and energy metabolism were altered upon application of the drug [90]. These kinds of studies have a great potential to

aid our understanding of insecticide drug action, and identify new drug targets in disease vectors.

One of the main advantages of *Drosophila* in biomedical research is the speed, at which drugs can be discovered and tested, which is much greater than in other model organisms (e.g. mouse), as well as low cost [91]. Due to fast life cycle, as well as visible and easily assessed phenotypes, many chemicals can be tested in the fly with high efficiency. Primary screens as well as secondary validation of potential drugs for various human disorders have been conducted in *Drosophila* [91]. For example, *Drosophila* model of fragile X syndrome (FXS) has been successfully developed, exhibiting phenotypes typical of FXS [92, 93]. The fruit fly was then used to screen a drug library of 2000 FDA approved potential therapeutics for FXS, 61 of which were discovered to rescue lethality in FXS flies, at least to a certain degree [94]. Subsequently, 15 of these compounds were chosen for further validation, and nine of these exhibited a dose-dependent effect for lethality rescue in FXS flies. These compounds require further validation in a mammalian model and subsequent human clinical trials to be approved as a treatment for FXS. However, primary screening in *Drosophila* can narrow down collections of potential therapeutic in a time- and cost-effective manner and hence greatly reduced the time and cost of the drug development process. Other than primary screening, flies can be used for a secondary validation screen [91]. For example, using *Drosophila* model of Huntington's disease has assisted development of new therapeutics that can inhibit polyglutamine protein-mediated aggregation [95]. Potential drugs were first screened using yeast as a primary platform. Nine selected compounds were then tested in mammalian cell culture models, identifying four compounds that were subsequently tested using mouse hippocampal slice culture [96]. One compound was then identified as the most effective in its ability to inhibit aggregation, and tested in a whole-animal model, *Drosophila* Huntington's disease model.

In summary, *Drosophila* models can be widely used in biomedicine, in order to aid understanding of genetic and molecular basis of pathological processes, as well as drug development and screening. The fly provides a relatively easy, time- and cost-effective platform for biomedical research, and has proven to be a successful tool for new discoveries. Moreover, it is particularly useful for studying rare diseases, for which there is limited funding, such as inborn errors of metabolism (IEMs) discussed in the next

section.

### 1.2.3.1 IEMs and *Drosophila*

Inborn errors of metabolism (IEMs) result from mutations in metabolic enzymes and comprise the majority of human genetic diseases. IEMs lead to accumulation of metabolites [97], which can be directly harmful (e.g. accumulation of upstream metabolites to toxic levels, or deficiency in downstream metabolites), or impose a burden on the urinary system leading to kidney damage [98]. Lack of feasible therapies as well as relatively little information on the molecular basis of IEMs necessitates the urgent development of new research approaches. However, a small number of IEM cases in Caucasian population limits research funding. Hence, development of a simple and cheap animal model is required for the study of IEMs and development of possible treatments. Importantly, the maintenance of fly lines is inexpensive and relatively easy. For the study of human metabolism and metabolic imbalances, fruit flies are particularly suitable for tissue-specific analysis and thus are an attractive *in vivo* 'bridge' between the unicellular organisms in which systems biology was developed, and the complexity of the human body.

The work in this thesis has been based on research on *Drosophila melanogaster* purine metabolism. Enzyme defects within purine metabolism result in IEMs including, xanthinuria type I [62, 63, 99] and type II [100, 101], and hyperuricemia [97, 102]. IEMs that have been successfully modelled in *Drosophila* include Xanthinuria type I (deficiency in *rosy*) [62, 63, 99] and type II (deficiency in *maroon like*) [100, 101]. Both mutations result in bloated MTs and sensitivity to dietary purines [103]. Both genes are involved in purine metabolism and the molecular basis of both IEMs has been elucidated thanks to studies in the fly [104-106]. Another IEM, which has a potential as a *Drosophila* model, is hyperuricemia, which is associated with many different disorders, including cardiovascular disorders (for example, coronary heart disease, congestive heart failure, stroke, and peripheral artery disease) [107-109], renal failure [98], hypertension [107-109], chronic kidney disease [107], metabolic syndrome (including insulin resistance) [110], and gout [111-117]. Humans are susceptible to hyperuricemia due to their inability to degrade urate [118, 119] (for details see Chapter 3). As a result urate levels in the blood are increased, which leads to oxidative stress and related conditions mentioned above. Fruit flies, on the other hand, break down urate to allantoin due to presence of



functional urate degradation enzymes in their genome (see Chapter 3). It would be interesting to establish *Drosophila* model of hyperuricemia in order to elucidate molecular basis of the disease and develop/test new therapeutics for this disease. Currently, there are no effective drugs for hyperuricemia, and the treatment is based on application of compounds to lower urate levels, such as urate oxidase agents and xanthine oxidase inhibitors (to lower the production of urate). These treatments are not very effective and result in numerous side effects [117, 120]. Hence, an animal model of the disease would allow identification of new potential therapeutics, as well as screening tool. Moreover, relatively little information on the molecular basis of IEMs necessitates the urgent development of new research approaches.

#### **1.2.3.2 The need to study individual tissues**

Previously, Dow/Davies lab has generated the most complete tissue-specific transcriptomic atlas for *Drosophila* – FlyAtlas.org, which provides data on gene expression across multiple tissues of adult fly and larva [6]. FlyAtlas revealed that transcripts of many genes are enriched in specific *Drosophila* tissues, and that it is possible to deduce the functions of individual tissues within the fly. Based on FlyAtlas data, it has become clear that the fly (like other metazoan species) must be considered as a set of tissues, each with its own distinct transcriptional and functional profile. Moreover, it revealed that for about 30% of the genome, reverse genetic methods (i.e. mutation in an unknown gene followed by observation of phenotype) are only useful if specific tissues are investigated. This stresses the importance of a research approach focusing on individual tissues as well as the whole organism, rather than the whole organisms only. Moreover, FlyAtlas showed that many fly homologues of human disease genes are enriched in fly tissues that have the same role of the equivalent human tissue [6].

Interestingly, the importance of tissue-specific metabolomics approach in other organisms have been noted, for example, in the understanding of the resistance against herbivores in *Jacobaea vulgaris* or metabolism of auxin in different tissues of *Arabidopsis thaliana* [121, 122].

Altogether, these findings show that it is important to consider an organism as a whole as well as a set of individual tissues in order to improve our understanding of organismal function, and validate *Drosophila* as a model organism.

### 1.3 Metabolomics

Metabolomics is one of the omics among genomics, transcriptomics and proteomics. The aim of metabolomics is to identify and quantify all small molecules (<1500 Da) in a biological system [123, 124]. All small molecules present in a sample are referred to as the metabolome [123, 124]. Metabolomics is unique in comparison to other omics in that it has a direct correlation to the physiology of the cell/tissue and reflects its activities at a functional level. As a result, it provides insights into complex phenotypes of biological systems [124, 125]. In contrast, other omics reflect the flow of gene expression but are not directly correlated to function and phenotype [125, 126]. Moreover, the diversity and dynamics of metabolites is much greater than these of genes or proteins.

Metabolomics is divided into two main approaches: targeted and untargeted approach. Targeted metabolomics focuses on a specific group of metabolites, for example, involved in the same enzymatic pathway or metabolites that belong to the same class of compounds. Most commonly its aim is to provide absolute quantification of metabolites of interest, which is necessary to elucidate metabolic dynamics of enzymes [125]. Untargeted (or non-targeted) metabolomics, on the other hand, is a comprehensive analysis of endogenous metabolites and metabolic shifts in response to different stimuli, including genetic mutation, environmental changes, disease, etc. [127-131]. It aims to simultaneously identify and relatively quantify all metabolites in a given system. It is a hypothesis-free approach, which generates new hypotheses. Subsequently, these can be tested by targeted profiling to provide absolute quantification of selected metabolites [127, 132].

Targeted metabolomics is typically performed using mass spectrometry (MS) based methods, including GC-MS (gas chromatography-MS), LC-MS/MS (liquid chromatography-MS/MS) or FIA-MS/MS (flow injection analysis MS/MS) [124, 133]. Absolute quantification is possible due to presence of authentic standards for each metabolite of interest, which are at known concentrations and are analysed in the same experiment. On the other hand, untargeted metabolomics typically employs MS or NMR (nuclear magnetic resonance) spectroscopy in order to measure relative abundance of all metabolites in the system [124]. As opposed to the targeted approach, absolute compound concentrations are not available, due to limited access to internal standards. However, a small set of standards is used to monitor the performance of analytical

instrument, and to provide semi-quantitative measurements for each metabolite.

### **1.3.1 What is metabolomics used for?**

Metabolomics has gained popularity in recent years and offers a huge potential to life science research [134]. There are plenty of advantages of holistic metabolomic profiling, which were described by Gika *et al.* [135]. In summary, by elucidating the metabolome of any given system, its real-time metabolic status is obtained, which reflects responses of the system to the environment it is in, including genetic perturbations, environmental stimuli, drugs, disease factors etc. Furthermore, untargeted metabolomics is a useful tool in the identification of novel markers (including markers of disease), which provide new insights into biochemical functions. Finally, metabolomics complements genomic, transcriptomics and proteomic data, which altogether aid the development of holistic understanding of the organismal function.

Metabolomics has been used in numerous research fields including medicine, animal biology, microbial research, food science [125, 127, 136], plant biology, including genetically modified (GM) crops [137] and research on the model plant *Arabidopsis thaliana* [138], environmental studies [139], functional genomics [140] and integrative systems biology [141-144].

#### **1.3.1.1 Medicine**

In biomedical research, metabolomics has been used to elucidate the correlation between metabolic changes and to understand physiological and pathophysiological processes [127]. Previous research has applied metabolomics to the study of kidney diseases [129] cardiac research [145], metabolic syndromes and related diseases, including cardiovascular disease [146], diabetes [147, 148], atherosclerosis [149] and ischemia [150].

Metabolomics is a powerful tool for biomarkers discovery [151], especially in case of diseases with complex diagnostic criteria, for example, psychiatric disease, including Parkinson's disease [152], Huntington disease [153] and schizophrenia [154]. Biomarkers have been discovered in renal diseases, including autosomal dominant polycystic kidney disease, kidney cancer and acute kidney injury [129]. Biomarkers can be also used to study toxicity and therapeutic effects. One of the biggest targets of medical metabolomics is cancer. Metabolomics tools can be used for its early diagnosis to

improve prognosis [125, 155-157].

Finally, metabolomics can improve the understanding of the mechanisms of drug action and the response of organisms to drugs [136, 158].

#### **1.3.1.2 Microbial research**

Metabolomics has been applied in microbial research, including studies in *Trypanosomes* [141, 142, 159], *Escherichia coli* [160] and *Saccharomyces* strains [161]. Various analytical approaches have been used, including snapshot metabolomics for phenotyping of silent mutations to aid the understanding of gene function [162], quantitative metabolite analysis by stable isotope dilution [163] and the relationship between cellular metabolome and global regulation, such as growth rate [164].

#### **1.3.1.3 Food science**

In food science, metabolomics has been used to evaluate quality and safety [125, 165]. Major crops have been studied, including potato [166] and fruits [167] as well as popular beverages: wine [168] and beer [169]. Metabolomics has also been applied in herbal medicine [125].

#### **1.3.1.4 Animal research**

Most of animal metabolomics focuses on physiology, development and disease of model organisms, including zebra fish [170-172], *Drosophila melanogaster* [173-177] (discussed in Chapter 5) and *Caenorhabditis elegans* [178, 179]. Metabolomics research in these organisms has provided new insights into development, biochemical profiles of different tissues, study of silent mutations, and metabolic changes triggered by different stimuli, including environmental stress, genetic perturbations and breeding. Altogether, it has provided a better understanding of organismal physiology, development and disease, which can be applied to research in other species, including humans. Moreover, accumulation and integration of metabolomics data in model organisms will provide a holistic understanding of organismal function.

### **1.3.2 Metabolomics technologies**

At present there is no single platform that can measure all metabolites in a biological system due to huge diversity of small molecules in terms of chemical structure and concentration as well as the dynamic nature of the metabolome [124, 127, 158]. Different

technologies have their advantages and limitations and are suitable for different type of analyses. It is important to take this into consideration when designing an experiment and choosing an appropriate instrument. It is a compromise between chemical selectivity, sensitivity, speed and cost. Moreover, the choice of technology often depends on existing collaborations and available tools.

The most commonly used technologies for metabolomics are NMR [128, 180-182] and MS - based methods [62, 106, 127, 177, 183-187]. For some analyses, which do not require high sensitivity but rely on higher throughput, other platforms can be employed, including Raman spectroscopy and Fourier transform infra-red (FTI) spectroscopy [188]. Other less commonly used separation methods include (CE) and supercritical fluid chromatography (SFC) [189]. A comparison between NMR and MS-based platforms for metabolomics is summarized in Table 1-1.

#### **1.3.2.1 NMR**

As illustrated in Table 1-1, NMR has many advantages and disadvantages in comparison to MS-based technology [136, 145, 190-192]. It has been commonly used in metabolomics research, including biomedical and pharmaceutical research [149, 181, 190] as well as drug discovery and toxicity [180, 182]. Developments in NMR techniques, including high-resolution  $^1\text{H}$  NMR spectroscopy and  $^{13}\text{C}$  NMR spectroscopy, made it a powerful tool for studying biofluids and intact tissues [193]. These technologies can be used to measure biochemical changes without separation and sample preparation. However, as is the case with other metabolomics platforms, obtained spectra require laborious analysis and data interpretations (see Chapter 1.3.3.3).

**Table 1-1 Comparison between MS-based and NMR metabolomics including advantages and disadvantages of both methods.**

Technology	Advantages	Disadvantages
MS	greater sensitivity greater selectivity precision robustness high resolution wide dynamic range coverage of a wide chemical diversity feasible elucidation of the molecular weight and structure of unknown compounds	variable detection responses, such as differential volatilization or ionization cost samples not reusable
NMR	high-information content of the spectra relative stability of NMR-chemical shifts ease of quantification no need to pre-select the conditions for the analysis cost samples can be reused no requirement for sample preparation, such as derivatization	higher amounts of sample required less sensitive, only detects most abundant metabolites lower resolution lower coverage of chemically diverse species

### 1.3.2.2 MS-based platforms

The overall sensitivity and resolution of MS is considered greater than that of NMR. Moreover, it covers a wider range of chemical diversity and is better suited for the identification of unknown metabolites [136, 194]. Many mass spectrometry instruments have been developed over the years and are described in [195]. In summary, different tools have advantages and limitations and require different sample preparation and analysis. There is a compromise between the number and the quality of detected ions by different MS systems [132]. One of the most commonly used platforms is time of flight

(TOF) MS. Its main advantages are high resolution, sensitivity and speed [196]. High resolution MS, such as Orbitrap MS is considered the most sensitive MS system [136] due to its high resolving power. This is achieved by decreasing the background noise and the interference from co-eluted compounds [197]. It allows unambiguous assignment of a molecular formula to many observed masses due to high mass accuracy, resolution and dynamic range [159]. Orbitrap has become more popular for the study of metabolomics in recent years [106, 186, 198]. In global metabolomics studies, it performs best when coupled to LC to analyse samples in both positive and negative ion modes. The Orbitrap Exactive tool is suitable for this kind of analysis due to its positive-negative polarity switch mode. For higher scan speed and higher quality metabolomics data, the Q-Exactive MS (QEMS) Orbitrap system is the best choice. It can be used simultaneously for targeted and untargeted metabolomics [199].

MS enables metabolite identification based on its mass, and compound quantification [127]. The coupling of MS to an additional separation technique allows higher quantitative accuracy and more confident metabolite identification. MS can be combined with the following systems: liquid chromatography (LC), high performance liquid chromatography (HPLC), ultra high performance liquid chromatography (UHPLC), gas chromatography (GC), capillary electrophoresis (CE), and flow injection analysis (FIA) [158]. Each separation method has its advantages and disadvantages and hence is most suited for different purposes. Additional separation reduces background noise, which results in higher quality MS spectra [127].

In summary, FIA is a fast analysis mostly applied in targeted metabolomics for simultaneous identification and quantification of multiple metabolites, where mass spectra for different compounds are available before analysis [158]. CE is another fast method, which results in separation of polar metabolites by charge, frictional forces and hydrodynamic radius [200]. It employs a capillary with high voltage, which allows electrophoretic migration of metabolites in an electro-osmotic flow [158]. GC separation has been commonly used to separate fatty acids and sugars [201]. This method can only be used to analyse volatile compounds or compounds that can be made volatile by derivatisation [202, 203] and hence is mainly limited by molecule size and volatility. GC-MS utilises electron ionisation (EI) to ionise and fragment volatile molecules. It requires more complicated sample preparation than other separation methods, for example, LC. Additionally, it is not suitable for detection of many polar compounds and can only

measure the most abundant positively charged ions because of the EI technique used in this method [204]. On the other hand, metabolite identification is much easier in GC-MS in comparison with LC-MS. This is because the fragments generated by EI are more reproducible and can be matched to existing GC-MS spectral databases. Moreover retention times produced by GC-MS are more stable compared to LC-MS [188].

### **1.3.2.3 LC-MS for metabolomics**

LC-MS (as well as FIA-MS) is one of the best methods for the detection of soluble and lipophilic metabolites [203, 205]. It has been commonly applied in proteomics, lipidomics and metabolomics [62, 90, 106, 177, 186, 187, 193]. More sensitive techniques were developed from LC-MS, including HPLC and UHPLC. In this thesis LC-MS (specifically UHPLC) was used for metabolomics analysis. Hence, this technique will be discussed here in more detail.

LC is one of the most commonly used separation methods, coupled to MS. Combining LC with MS reduces samples complexity and separates metabolites before detection. It enables separation of isomers, which would otherwise appear as same metabolites in MS analysis, and reduces ion suppression of less ionisable compounds by more ionisable ones [206]. Dissolved compounds are separated in a liquid mobile phase along a solid stationary phase (for example, silica) [158]. As opposed to GC, it utilises electrospray ionization (ESI) for molecule ionisation and separation [207]. The separation phase is relatively longer than in case of CE and FIA. However, in UHPLC separation is increased and faster thanks to smaller particle size of material used for the stationary phase [208]. Moreover, HPLC can be used for separation of molecules of a wide range of polarity, which makes it a very versatile and powerful tool.

Different systems have been developed to perform LC, including reverse-phase (RP) columns, Sub-2 micron based columns and HILIC (hydrophilic interaction liquid chromatography) columns [135, 209-211]. RP LC provides high resolution, wide metabolite coverage and high reproducibility and is most commonly used for detection of non-polar and medium polar metabolites [132, 189]. Sub-2 micron based columns in combination with UHPLC are currently the most popular method for metabolic analysis of clinical samples [212, 213]. It provides higher throughput than other LC-MS techniques, with increased sensitivity and higher resolution [209]. One of the most popular LC systems, together with RP LC, is the HILIC column [186, 214, 215]. It has an advantage



over RP as it allows separation of polar and lipophilic metabolites. It enables retention of polar metabolites, while lipophilic metabolites elute rapidly from the column. On the other hand, lipids do not elute easily from the RP column, which can lead to their accumulation and ion suppression [106]. As a result, HILIC column is the preferred method for separation of polar and charged metabolites [186, 216].

Within the HILIC system, there are different types of columns available, for example, with a different inner diameter. Comparison of 2.1 mm and 4.6 mm columns has been described in [135]. On the 2.1 mm column more compounds were putatively identified [217]. However, the results obtained from the wider column were more reproducible. Hence, both columns are suitable for different applications, for example, 4.6 mm HILIC is more suitable for large batch analysis. There is no single LC-MS method that can analyse all types of compounds. In order to obtain the most comprehensive results, the combination of different types of columns might be the best solution. In order to improve the resolution of LC, new systems have been developed, including two-dimensional (2D) LC systems [218, 219]. These systems can provide better separation and coverage as well as higher resolution than other LC methods. However, further research is required to optimise these techniques.

LC-MS is considered a highly sensitive metabolomics method, much more so than NMR spectroscopy. It requires minimal sample preparation and enables effective ionisation with minimal fragmentation of a large variety of compounds. Hence, it offers a greater metabolite coverage than GC-MS [132]. Moreover, it can be coupled with tandem MS (MS/MS) to obtain compound structural information, which is not possible in case of CE-MS [183].

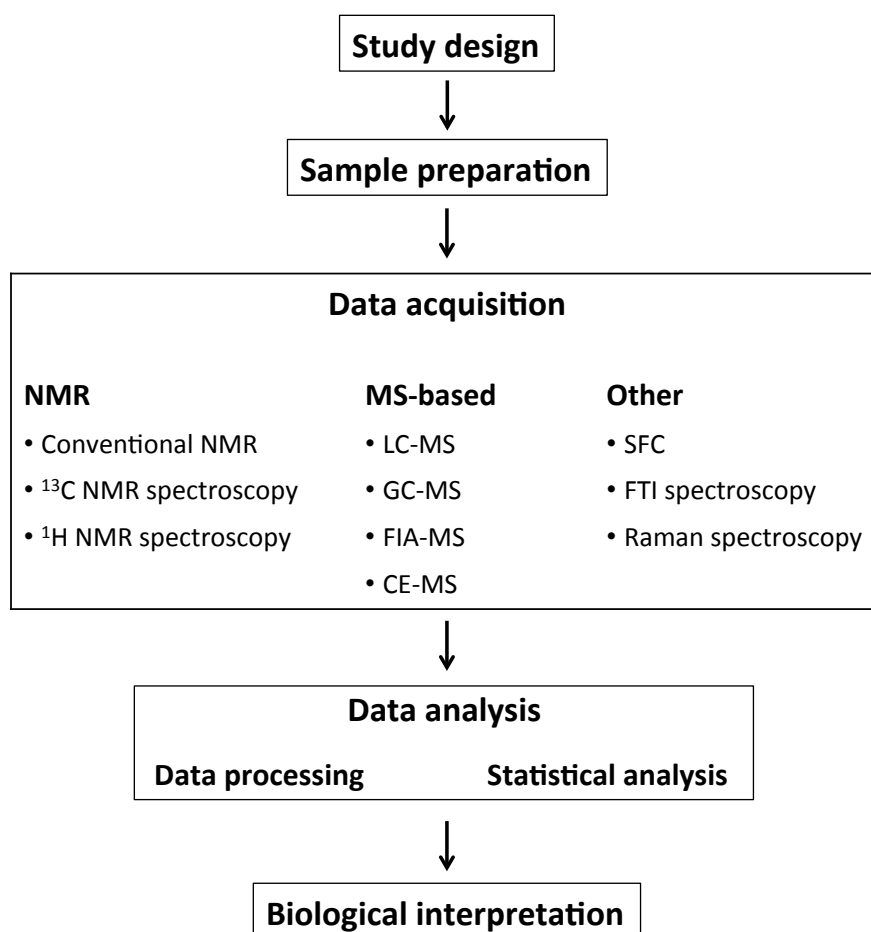
### **1.3.3 Sample preparation for LC-MS analysis**

Since metabolomics methods are high-resolution techniques, they require highly controlled and strict experimental conditions [125]. Regardless of a chosen LC-MS system, several steps must be followed in order to obtain high quality metabolomics data. These consist of experimental design, sample collection and pre-treatment, instrument analysis, data analysis and biological explanation [132, 220]. Metabolomics workflow is summarised in Figure 1-2.

Appropriate experimental design is crucial to obtaining high quality metabolomics data.

One of the considerations is selection of controls, which should be of the same genetic background and grown in the same strictly controlled conditions [135]. Moreover, quality control (QC) samples should be analysed with each experiment to ensure measurement repeatability. QC samples are usually prepared from a mixture of small aliquots of each sample in the experiment. Previous research showed that many factors could affect sample quality and the result of metabolomics analysis, including type of collection tubes [221] as well as sample storage and transportation [222]. Hence, it is critical for the experimental conditions to be strictly controlled, and sampling protocol to be standardised in order to minimise biological and technical variability [158]. Another consideration in terms of sample collection is homogenisation and quenching (bringing metabolism rapidly to a halt). This can be achieved by using liquid nitrogen for quenching, or organic solvents (for example, methanol, isopropanol, chloroform or their mixture) for quenching, metabolite extraction and protein precipitation [132]. Finally, chemical derivatisation is commonly required in case of LC-MS and GC-MS in order to improve compound ionisation before separation.

Following experimental design and appropriate sample preparation, LC-MS measurement is conducted. In summary, LC-MS provides exact mass measurements (from MS) and retention time (RT) values (from LC) that define ion features ( $m/z$ , RT). This is the first step towards compound identification. Different LC-MS platforms can be used as described in the previous chapter. Following primary data acquisition, data processing must be performed for metabolite identification and analysis [123].



**Fig. 1-2 Scheme of metabolomics workflow. Adapted from [136].**

#### **1.3.4 LC-MS data processing**

Data processing and biological interpretation of metabolomics data are the most laborious and challenging steps of metabolomics analysis, including LC-MS metabolomics. Acquired datasets are very complex and diverse. In order to enable metabolomics data processing, plenty of bioinformatics tools (utilising different programming languages, tools and algorithms) have been developed in recent years, for review see [130]. In this thesis, untargeted LC-MS-based metabolomics analysis was performed. Hence, here I will focus on data processing methods applied to global LC-MS metabolomics. However, many of described tools can also be successfully used for targeted metabolomics data.

In an untargeted LC-MS experiment, metabolome data of the whole sample are acquired. General processing steps are then followed, including data pre-treatment, peak detection [223], peak matching, and steps resulting in signal filtering and noise removal [224]. These steps allow putative identification of compounds based on exact mass and predicted retention time, or accurate identification, where authentic standards are

available. Following metabolite identification, multivariate and univariate data analyses are performed in order to elucidate statistically significant results. The final step of data processing is correlation of the results to metabolic databases and pathways, and biological explanation of the results [130, 132].

#### **1.3.4.1 Data pre-treatment**

Data pre-treatment and processing is the initial step after raw data acquisition. It commonly involves MS spectra normalisation, clustering and quality assessment, internal standard calibration, as well as precursor charge determination [225]. Data pre-treatment is necessary due to two-dimensional nature of LC-MS data [136]. As a result, it provides easier access to ion features, such as retention time,  $m/z$  values, ion intensity and isotope distribution [130].

#### **1.3.4.2 Peak annotation and matching**

Following data pre-processing, peak/feature detection and matching (also referred to as peak picking) is performed [158], and can be achieved using several different bioinformatics tools, including XCMS (several versions are currently available) [226-228], MzMine [229, 230], msInspect [231], MetAlign [232], CentWave [233], and many more [234]. Coble *et al.* reported that the best tools for peak detection for LC-MS data are MetAlign, XCMS, and MZmine [235]. Moreover, XCMS is an open-source algorithm and is the most commonly used for feature detection of LC-MS data [123].

One of major challenges in the processing of MS data are missing values, which can arise due to technical and biological reasons [130]. They are a result of metabolite absence or their abundance below detection limit. Missing values are particularly problematic for statistical analysis, which is not optimised to process 'zero' values. They can be dealt with by using numerous tools and data imputation, such as x-VAST [236], metabomxtr [237] and MzMatch [238, 239]. Another source of variance in LC-MS data is retention time drift. Retention time drifts are relatively common in LC-based systems. As a result two different features can be annotated as the same feature and then compared across different samples, leading to incorrect data interpretation. Once again, different tools have been developed to deal with RT drift [130]. Moreover, many peak picking tools, including XCMS, can carry out RT correction and features grouping to overcome this issue. As a result of peak picking, a list of all detected peaks is generated.

The data is then processed in order to match and group peaks of four biological replicates from each sample, to generate a single dataset for each sample. This can be achieved using parameters, such as PPM (mass error in parts per million, PPM), and RT window, which represent deviation from sample to sample in terms of mass and RT respectively. Moreover, noise reduction steps are performed, including RSD (relative standard deviation), noise, intensity and detections filters [238]. RSD filter elucidates peak reproducibility for replicates of each sample, noise filter assesses peak shape, whereas intensity filter removes features that are below intensity threshold in all samples, and detection filter determines a minimum number of samples, in which a feature peak has to be present. Furthermore, related peaks/ion annotation is performed to distinguish features, which originated from the same compound. This is required because LC-MS measurements often produce multiple peaks that correspond to the same metabolite. These peaks represent isotopes, dimers, adducts and fragments, which have different  $m/z$  values but similar retention times [127]. These can be recognised because their extracted ion chromatograms (EICs) have similar signal shape and intensity patterns [158], and then clustered together. Noise reduction methods and related peaks annotation can be performed using software tools, such as MzMatch [238, 239], CAMERA [240], PUTMEDID-LCMS [241].

#### **1.3.4.3 Metabolite identification**

The next step in LC-MS data processing is metabolite identification, which is the major bottleneck in LC-MS. It is based on neutral masses of each compound calculated from mass differences of adducts/isotopes [127, 238]. These calculated masses are then searched against existing metabolite databases, such as HMDB (Human Metabolome Database) [191, 242], Metlin [224] and KEGG (Kyoto Encyclopaedia of Genes and Genomes) (<http://www.genome.jp/kegg/>) [191, 224, 243]. KEGG is currently the most comprehensive knowledge base on organismal metabolism. Metabolites with molecular masses within the specified tolerance range to the query  $m/z$  value (within specified mass error (PPM)) are extracted from databases as putative identifications [127]. However, confident metabolite identification is rarely possible by matching calculated mass to the database, which commonly results in multiple identifications, due to presence of isomers (one feature can have ion can have over 100 putative identifications) and relative inaccuracy of MS [244]. To increase the efficiency of feature identification, additional methods have been developed. Creek *et al.* proposed a model called quantitative

structure retention relationship (QSRR) for elucidation of predicted retention times for compounds similar to authentic standards [238]. By adding this step, metabolite identification was highly improved, and 40% of falsely identified compound were removed. Moreover, for higher specificity features, preferential fragmentation can be performed, which provides additional fragmentation data. This allows more confident metabolite annotation [123]. However, in order to obtain reliable and unambiguous metabolite identification, mass, retention time and fragmentation spectrum must be compared to those of authentic standards analysed in the same LC-MS experiment [158].

#### **1.3.4.4 Statistical analysis of LC-MS data**

Once metabolite identifications are acquired, the data can be analysed using various statistical tools [130], in order to extract biologically relevant information [245]. This is a crucial step in obtaining meaningful metabolomics data before biological interpretation. Its aim is to determine the cause of unwanted variation (due to various biological and technical reasons, such as instrumental errors) and normalise the data by removing these variations [246]. The 'real' differences can then be assessed and analysed in order to obtain statistically significant results. Statistical analyses can be divided into two groups: univariate and multivariate methods. Univariate statistics includes methods, such as t-test, ANOVA (analysis of variance) and multiple regression, and analyses a single variable of interest. In metabolomics analysis, this means that these methods are used to analyse each feature independently, for example, the analysis is used to elucidate whether difference in metabolite abundance between samples are statistically significant. Multivariate statistics, on the other hand includes principal component analysis (PCA) [247], partial-least-squares discriminant analysis (PLS-DA) [248], hierarchical cluster analysis (HCA) [136], discriminate analysis, and self organizing networks (GEDI) [32]. These are used to analyse two to more variables of interest, for example, in case of PCA, all features of all samples are analysed resulting in simplified representation of the data, providing clustering tendency and outliers. PCA is recommended as a first step in metabolomics data analysis as it provides an overview of the patterns of each sample and relationships between different samples. These statistical tools can reduce the complexity of data by allowing researchers to focus efforts on groups of features that exhibit similar abundance patterns. Among most commonly used statistical tool for metabolomics data analysis are PCA, PLS-DA, the *t*-test and ANOVA [249]. In recent years, many statistical platforms were developed for metabolomics in order to make the data analysis more

varied, efficient and user-friendly (for review see [130, 158]). A commonly used, freely available statistical package that enables a wide variety of data analyses is MultiExperiment Viewer (MeV) [250]. It has been widely used for microarray data analysis, but also provides some very useful tools for metabolomics data analysis.

#### **1.3.4.5 Biological interpretation of LC-MS data**

The final step in metabolomics data processing is biological interpretation. One of the most important steps towards it is linking the obtained results to known metabolic pathways. This can be achieved by pathway analysis, network reconstruction and visualisation tools [158]. This kind of analysis for global metabolome is only possible thanks to bioinformatics tool. Manual analysis would be extremely time consuming and expensive. There are plenty of bioinformatics tools available and some examples include TrackSM, which assigns chemical compounds, based on their molecular structure, to metabolic pathways [251]; MarVis-Pathway enables annotation of pathways from different omics data, and provides pathway enrichment and meta-analysis [252]; PathWhiz is a server that interprets metabolomics data by assigning metabolites to known pathways, and generates colourful and aesthetic results [253]; InCroMAP is used for data integration, analysis and visualisation and provides enrichment analysis and pathway visualisation. It also deals with multiomics data [254].

#### **1.3.4.6 Tools for metabolomics data processing**

Recent years witnessed a huge development of bioinformatics tools for metabolomics data analysis. These include tools for data pre-treatment, peak picking, noise removal, statistical analysis and pathway analysis etc., as well as multifunctional tools that can deal with raw LC-MS data and perform all the necessary sample processing, annotation and statistical analysis [158]. For a comprehensive review of newly developed bioinformatics tools for LC-MS data processing and analysis see [158]. Among available software for the processing of raw LC-MS data, one of the most commonly used toolkits is mzMatch [255]. It includes noise filtering, gap filling, peak matching, related peaks annotation and compound identification by matching calculated masses to databases [239]. Moreover, it uses PeakML file format, which is compatible with other software, including XCMS, mzMine and IDEOM [238]. This allows flexible access to additional data processing tools. For example, IDEOM is an Excel interface to mzMatch, which allows comprehensive data processing of LC-MS data followed by statistical analysis and data visualisation tools

[238].

In terms of statistical analysis, data processing software include BioStatFlow version 2.7.7. (<http://www.inra.fr/>), which is a web tool for analysis of –omics data, including metabolomics. Another tool, RepExplore, is also a web service, which provides fully automated data processing as well as statistical tools, including ranking tables, whisker plots, heat maps, and PCA [256]. Another open-source tool, Normalyzer, can be employed for data normalisation using 12 different methods [257]. Multivariate statistical analyses can be performed using platforms, such as DeviumWeb (Dynamic Multivariate Data Analysis and Visualization Platform) (<https://github.com/dgrapov/DeviumWeb>). It provides statistical analysis, such as data normalization, clustering, PCA etc., as well as data interpretation (pathway enrichment analysis) and visualization.

Among multifunctional tools, the most commonly used ones are MetaboAnalyst ([www.metaboanalyst.ca](http://www.metaboanalyst.ca)) [258], XCMS Online (<http://masspec.scripps.edu/>) [227], MASSyPup [259], IDEOM [238], and MeKO [260]. They are all designed to analyse raw metabolomics data, perform all the necessary steps for data processing, annotation, data analysis, interpretation and visualisation. They require minimal bioinformatics knowledge, minimise the time spent on data processing and provide multiple statistical and visualisation tools. However, it is important to adjust the settings of these tools accordingly to the analysis performed to avoid user bias [158]. For a comprehensive review of analytical tools for metabolomics from data processing to pathway analysis, see [261].



### **1.3.5 Limitations and challenges of LC-MS for metabolomics**

Despite recent developments in metabolomics techniques and tools, metabolomics analysis presents numerous challenges due to technological limitations as well as issues with data interpretation [158]. Metabolomics measurements commonly identify more than 40,000 features, which have diverse physical and chemical properties [136, 191, 262]. Moreover, the range of compound concentrations is very broad, with over 12 orders of magnitude [132]. As a result of this complexity, and due to the dynamic nature of a metabolome, there is no single platform that can detect and analyse all metabolites in a given system [124, 127, 130, 158].

#### **1.3.5.1 Metabolite identification**

The main challenge in MS data analysis, prior to biological interpretation, is metabolite identification [127]. It has been estimated that about 70% of features extracted from LC-MS are not used in the final data analysis due to lack of accurate identification. As explained in previous chapter, metabolite annotation based on detected mass of compounds gives rise to putative identification, which can result in multiple identifications, due to presence of isomers and relative inaccuracy of MS [244]. Combining mass data with predicted and standards-matched retention times enables more accurate identification. However, in order to obtain confident metabolite identification, internal standards for every metabolite has to be analysed in the same experiment. Due to large size of untargeted metabolomics datasets, this is not feasible. This is because standards are not available commercially for all analysed compounds, and a lot of metabolites are not stable and hence obtaining standards is impossible [135]. An exception, where no standards are required, is ultra high performance liquid chromatography coupled with Fourier transform ion cyclotron resonance mass spectrometry (UHPLC–FT- ICR–MS) [208], which provides an accurate metabolite mass with a great precision allowing evaluation of its molecular composition. This enables confident metabolite identification without addition of authentic standards [263].

One of the main limitations of LC-MS is the fact that it only enables semi-quantitative analysis [47]. Challenges in metabolite identification and quantification arise due to technical issues in LC-MS analysis because LC-MS spectra are highly variable [123]. This is due to instrumental drifts, for example retention time drifts [264] or MS detection instability [135]. Moreover, instrument sensitivity presents a limitation in case of poorly

ionised metabolites or compounds at a very low abundance [127]. As mentioned earlier, metabolites can be chemically derivatised in order to improved ionisation. However, derivatisation does not always solve this problem. A novel approach, called isotope coded derivatisation (ICD) further enhances ionisation, and addition of isotope labelled derivatisation reagents results in the generation of isotope labelled ions as well as non-labelled ones [265-267]. These methods improve feature separation and identification, and thanks to isotope labelled standards, absolute quantification of selected metabolites is possible [28,60] in LC-MS analysis. Moreover, technical improvements could improve the reproducibility and resolution of LC-MS spectra, for example designing more stable columns, mass spectrometers with higher resolving power and mass accuracy etc. [158].

One of the strategies that can improve metabolite identification is preferential fragmentation achieved by MS/MS, which provides additional fragmentation data and chemical structure [268, 269]. The fragmentation spectra can be then compared to spectra of tandem MS databases, such as MassBank [<http://www.massbank.jp>, [223]], METLIN [<http://metlin.scripps.edu/>, [224], or NIST MS/MS [<http://www.nist.gov/mml/bmd/data/tandemmass-spec.lib.cfm>]. However, in contrast to well-annotated gene and protein databases, these databases are not complete. However, they can aid metabolite identification by elucidating its structure or determine its function. Moreover, fragmentation data are highly depend on the type of LC-MS instrument, and hence they are not highly replicable [135]. As a result, there are no comprehensive spectral libraries for LC-MS, which could be used for confident metabolite identification [158]. Hence, when standards are not available, spectral data cannot be evaluated using general databases for mass fragmentation spectra. This is the main advantage of GC-MS over LC-MS. In case of GC-MS, fragmentation spectra are stable and can be compared to existing spectral databases [136].

In order to improve the process of searching tandem MS databases, several tools have been developed, such as BioSM [270], competitive fragmentation modelling (CFM) [271], HAMMER [272] and many more. If a metabolite is not found in a database and a standard is not available, it is impossible to provide 100% identification.

#### **1.3.5.2 Data interpretation**

Another major challenge in metabolomics is data interpretation [158]. LC-MS analysis gives rise to large and complex datasets, which are impossible to interpret manually. As a result, specific software tools have been developed in order to enable this process, as described in previous chapter. An overview of currently available tools is presented here [234]. Data interpretation is the most time consuming and laborious step in metabolomics data analysis. It requires a lot of biochemical expertise as well as programming knowledge, in order to make the most out of available bioinformatics tools. Despite constant development of new platforms for pathway analysis, network reconstruction and visualisation tools [158] there is a requirement for more user-friendly software, with increased capacity to process complex datasets and better correlation of data to biological phenotype [130, 135]. Moreover, as is the case with other steps in metabolomics data analysis, standardised protocols for data interpretation are required in order to produce results with real biological relevance and that can be compared to other experiments. Another solution is to analyse the same datasets using different bioinformatics tools to provide the most reliable set of results.

#### **1.3.5.3 Data visualisation**

Once LC-MS data are analysed and interpreted, visualisation of the findings presents another challenge. Multiple complex metabolic pathways are difficult to represent in a simple and understandable format. Pathway analysis tools have been developed to aid this process, which take advantage of existing databases, such as KEGG [273] or HMDB [191, 242], including Biocyc (<http://biocyc.org>), Metabolights [274], Reactome (<http://www.reactome.org>), MGI Genome (<http://www.informatics.jax.org>), and MassTriX [275].

#### **1.3.5.4 Integration of metabolomics data**

Finally, challenges in LC-MS metabolomics arise from inability to compare data analysed in different laboratories and/or on different platforms [276], which is not the case for spectra obtained from NMR and GC-MS/MS experiments [191, 192]. This is particularly limiting in case of large-scale studies. Other than differences between LC-MS instruments, this is due to the lack of standardised protocols for experimental design, sample preparation and analysis [135]. Moreover, different data analysis platforms use

incompatible programming languages and are only available for single operating systems. In order to increase data reproducibility and the ability to correlate results from different experiments, standard protocols of metabolomics analysis need to be developed and followed by all researchers involved in metabolomics studies [123, 132]. Moreover, new platforms allowing normalisation methods to combine data collected at different times, using different instruments etc. are required.

Finally, in order to improve metabolite coverage and data quality, new metabolomics techniques have to be developed. At present, this issue can be tackled by using different chromatographic columns and improved compound ionisation [132, 277, 278]. Moreover, combining targeted and untargeted metabolomics approaches would yield more understanding of the tested system, including improved coverage and quantification [279, 280]. Moreover, in order to improve metabolite identification and structure prediction based on tandem MS, new LC/MS and MS/MS databases, as well as search algorithms are required [226].

In summary, major challenges exist in LC-MS data analysis, including metabolite coverage, compound identification and quantification, measurement reproducibility, data interpretation, visualisation and results correlation between different experiments. There are numerous strategies and tools available to improve these limitations. However, it is currently necessary to develop robust annotation methods, reliable metabolomics databases, and more sensitive and accurate instruments, improve data comparability, and expand current analytical platforms to provide more user-friendly multifunctional tools for data processing.

## **1.4 Systems biology**

Living systems are complex and this complexity is perhaps particularly noticeable when looking at vast datasets produced by high-throughput techniques, including metabolomics, genomics, transcriptomics etc. Systems biology approaches have been used to integrate this kind of data and obtain a holistic picture of organismal function. Systems biology is a field of science that studies whole living systems rather than its individual components in isolation. As opposed to the reductionist approach, which focuses for example on single genes or proteins, systems approach looks at networks and

interactions between all components of the whole system. Moreover, it attempts to understand biological processes in quantitative and qualitative way [281]. This is achieved by observing patterns of behaviour in a complex system and elucidating principles that govern it. These can then be applied to generate predictive models of biological systems [282]. The models can then provide information on how manipulating the system can affect it, for example a model of healthy and diseased states of the system [283], or they can be applied to other living systems.

Systems biology aims to condense the knowledge and asks questions in the context of a whole system. For example, in terms of single-gene mutation, it does not only look at its direct effect on protein expression, but also how it affects activity of other genes, proteins and metabolites and how they interact as a result of the genetic manipulation [284, 285]. Developing a model to test this kind of genetic challenges, allows a fast and often correct prediction of how the system will behave in different conditions. This can also be used in, for example, faster preliminary drug screening.

Systems biology is particularly useful for the study of multicellular organisms. This is because, for example, it is particularly important to take into consideration cross talk between different tissues or cell types in these organisms. It is useful to look at individual pathways or cells. However, in order to obtain a clear understanding of organismal function, we must look at the living system as a whole. For example, single gene mutation in *ry* has been shown to not only affect metabolites directly upstream and downstream of the lesion, but also metabolites several reactions away from it, as well as apparently unrelated metabolic pathways [62]. Moreover, in order to understand the basis of phenotypic manifestations caused by the mutation, it is necessary to take the whole system into consideration, rather than just a single component. It is clear from this study that single gene alterations have a much broader effect than altered expression of a single protein. It suggests that the results of genetic manipulation depend on numerous factors, including, cell type, developmental stage, and cellular processes, in which it functions. Moreover, compensatory mechanisms exist that can overcome genetic manipulations and affect the phenotype [286].

#### **1.4.1 Functional genomics for systems biology**

Functional genomics technologies provide vast wealth of data produced by 'omics'

methods, including genomics, transcriptomics, proteomics and metabolomics. They provide semi-quantitative data on the levels of genes, proteins and metabolites in the living system [287]. This enables to observe effects of conditional perturbation on an organism, including the effects on interactions between macromolecules, functional correlations between phenotypes, and the dynamics of the system [288]. The goal of functional genomics is to experimentally validate gene function and to elucidate the link between the functional phenotype and genes, between DNA sequence and function [162, 289, 290].

The ‘omics’ data are the key to constructing biological models at different stages and different conditions of the living system, for example at different environmental conditions or disease states [291]. This is exactly the aim of systems biology, to generate mathematical models and use them to predict a response of a living system to different stimuli. Altogether, this provides a more holistic understanding of the behaviour of a complex system.

Finally, thanks to a wider use of ‘omics’ technologies in recent years, hypothesis generation and testing has been accelerated. For example, development of FlyAtlas.org, tissue-specific transcriptome of *Drosophila melanogaster*, allowed prediction of tissue-specific activities and functions in the fly [6].

#### **1.4.2 Metabolomics for functional genomics**

Metabolomics is one of the techniques employed in functional genomics. It forms the basis for understanding the biochemical phenotype of an organism, and aids understanding of gene function by analysis of the metabolome [288]. As opposed to other omics, metabolomics has a direct correlation to the physiology of the cell/tissue, provides the most functional information, and hence is directly linked to phenotype [124, 125, 142]. In contrast, other omics reflect the flow of gene expression but are not directly correlated to function and phenotype [125, 126, 292]. Metabolomics aids the understanding of real functional correlations between metabolite changes and physiological/ developmental phenotypes [293]. These can then be linked to the gene(s) responsible for the metabolite change, and as a result elucidate gene function.

A large proportion of genes present in the genome encode protein of unknown function. Among these genes, many are ‘silent’, which means that they do not produce any

apparent phenotype when they are disrupted [294]. Metabolomics has been employed to elucidate the function of these genes [162]. It provides identification and relative quantification of all compounds in the metabolic network. By comparing metabolite concentrations to the concentration of one selected metabolite, the site of action of a silent gene can be elucidated. Moreover, metabolomics can provide metabolic profiles in mutants of unknown genes. By comparing these to profiles of controls with known genetic backgrounds, metabolomics can reveal the function of unknown genes [162].

Altogether, metabolomics is a powerful tool in functional genomics, used to predict and define complex phenotypes in biological systems [125]. It complements other omics, which reflect the flow of gene expression, and is directly linked to cell phenotype, physiology and function [126]. As a result, metabolomics has the potential to elucidate gene function of poorly characterised genes, and map phenotypes of silent mutations [295, 296].

### **1.4.3 Metabolomics – a systems biology tool**

Metabolomics provides an insight into the mechanisms underlying biological and biochemical processes of living systems [127]. It simultaneously identifies and quantifies all small molecules, and provides a comprehensive understanding of all biochemical reactions and metabolite fluxes in a system [297]. Hence, it is the core of systems biology, which aims to elucidate obtain a holistic picture of organismal function [298]. Moreover, combining metabolomics with genomics, transcriptomics and proteomics data can provide a unique insight into the interactions between different components of a living system [158]. Integration of different omics data can be used for the construction of molecular networks, which can aid the understanding of the complex biochemical processes in a system [299]. Moreover, it can reveal how these complex interaction lead to a specific behaviour and result in a phenotype [300]. This enables capturing the full functional genomic picture of an organism.

As opposed to functional genomics, systems biology does not study individual genes, proteins, or metabolites. On the other hand, it looks at interactions and relationships of all elements in a given system, while it is functioning [301]. The general approach is to disturb the system (by genetic manipulation, environmental change etc.) and then

observe the effects of the perturbation at the genomic, proteomic, and metabolomic level [126]. Obtained omics data can then be combined and used to generate computational models of organismal function. This method has a great potential in metabolite engineering, drug discovery, and medicine, for diagnosis and prevention of disease [302].

However, the integration of metabolomics data with other omics data poses several challenges. Data obtained from different platforms have different formats, and samples are collected using different experimental conditions, which altogether might be difficult to correlate. Moreover, the datasets are very large and complex, and require improved bioinformatics tools to process and interpret [158].

Current platforms for omics data integration include KeyPathwayMiner 4.0 [303], MarVis-Pathway, which enables annotation of pathways from different omics data [252], InCroMAP, which can be used for multiomics data integration, analysis and visualisation and provides enrichment analysis and pathway visualization [254].

#### **1.4.4 Genome-scale metabolic map reconstruction**

As mentioned earlier, one of the main goals of systems biology is to elucidate principles that govern a living systems, and apply them to generate predictive mathematical models of the system function [282]. Genome-scale metabolic map reconstructions or genome-scale metabolic models (GEMs) are an example of this kind of models. They are a huge step up from one-dimensional biochemical databases, such as KEGG [158]. GEMs are mathematical and computational models containing data on all metabolites, enzyme-coding genes, and biochemical reactions in a given system (usually a specific cell or whole organism) [304, 305]. Moreover, they provide representation of cellular biochemistry, and information on biophysical constraints on the system, such as oxygen and nutrient availability [306]. They facilitate the study of metabolic pathways in the context of the whole genome [307]. Furthermore, GEMs can be compared to experimental data to increase their validity [307].

These computational models can be studied using various mathematical algorithms, in order to answer important biological questions [308]. For example, they can be used to predict phenotypes by changing environmental conditions or introducing genetic mutations [309]. This have been used to model physiological and pathological states of



living systems, for example disease [283, 305]. Human global metabolic map Recon 1 has been applied to study cancer, diabetes, host-pathogen interactions, and heritable metabolic disorders [310-313]. Another purpose of GEMs is to generate and test hypotheses [304, 314]. For example, by introducing genetic mutations of poorly characterised or unknown genes, their effect on the system can be observed, and as a result their function predicted. These observations combined with experimental data provide a powerful tool for functional genomics. Finally, GEMs can be used to identify biomarkers and novel drug targets [305]. For example, a global metabolic models of *Mycobacterium tuberculosis* [315], *Campylobacter jejuni* [316], and *Plasmodium falciparum* [317], have been utilised towards development of new therapeutics.

In recent years, over 70 genome-scale metabolic map reconstructions have been generated of various organisms [306, 318], including *Plasmodium falciparum* [314], *E. coli* [319], *Mycobacterium tuberculosis* [315], *Campylobacter jejuni* [316], *Pseudomonas aeruginosa*, *Pseudomonas putida* [320, 321], human [310] and mouse [322]. These models are available a publically available knowledgebase of Biochemically, Genetically and Genomically (BiGG) structured genome-scale metabolic network reconstructions (<http://bigg.ucsd.edu>) [304]. The website provides access to 75 standardised GEMs, which are connected to annotations and external databases.

Stoichiometric metabolic maps are built by integrating the genome sequence and annotation of an organism with many different resources on organismal metabolism, including databases (e.g. KEGG [323], BRENDA [324], BioCyc [325], MetaCyc [325]) and primary literature [326]. Maps can be generated using semi-automated tools [158, 305]. A review of available tools is described in [158]. Some examples include PathCase Metabolomics Analysis Workbench (PathCaseMAW), used for mammalian map reconstruction [327], MetaMapR, which is based on KEGG and PubChem [328], and Integrated Interactome System (IIS) [329]. In constraint-based models, constraints can be then applied (biophysical constraints on reaction and transport rates), followed by expressing each reaction mathematically. Finally, different model calculations and predictions can be performed using computational mathematical tools, such as the COBRA toolbox for MATLAB [330, 331]. Generated GEMs contain missing information, where there is no connectivity between metabolites in the model, and as a result it is incomplete and does not produce biomass [311]. This indicates missing reactions and/or

pathways, and has to be manually curated with the aid of sequence alignments and literature to allow connectivity within the model [319].

#### **1.4.4.1 Gaps in GEMs**

All GEMs contain gaps, where the model produces biomass effectively but reactions are blocked and do not carry flux at steady state because metabolites are either only produced or consumed [332-334]. These gaps are a result of limitations of the model or knowledge gap, where our knowledge on organismal metabolism is limited. Different types of gaps are present in metabolic maps, where enzymatic reaction is absent [335]:

1. The homologue of enzyme-coding gene is present in an organism of interest but there is no experimental evidence confirming the function of the homologous gene
2. Enzyme exists in other species but there is no homologue in the genome of the organism of interest
3. Global gap (orphan enzyme), where no sequence of enzyme-coding gene is available in any known species

The first type of gap is the easiest to fill in, and requires integration of experimental evidence with the map. However, in practice, experimental evidence is not always easy to obtain due to complexity and dynamic nature of metabolism. The second type of gap can result from various scenarios. For example, the reaction occurs spontaneously in the organism of interest or is catalysed by an enzyme unique to this species or a group of species. Another possibility is that there is an alternative pathway producing the metabolite, or a symbiotic organism (e.g. gut microbiota) that carries out the reaction. Finally, the whole pathway may not be required in the organism of interest, for example bacteria-specific pathway in the fruit fly. The third type of gap is the most challenging one. Orphan enzymes have been predicted to represent 30-40% of all known enzyme activities [336-338]. It is particularly difficult to predict function of these genes. However, many of these genes are thought to be of ancient origin, and involved in essential metabolic reactions [337, 339]. GEMs present a tool to identify and study these gaps in the context of available metabolic knowledge, and hence are a powerful tool for gene discovery, and understanding essential metabolic processes.

#### **1.4.4.2 Filling in the gaps**

Several computations tools and algorithms have been developed to identify network gaps and aid gap filling process [335]. For constraint-based models several methods exist,

including SMILEY [340], GapFind/GapFill [341] and GrowMatch [342]. SMILEY has been successfully used to predict gap-filling reactions in the human GEM (Recon 1) [311]. The other two methods have been used in predictions of *E. coli* and yeast metabolic gaps [342]. For non-constraint-based models, other algorithms can be employed, such as Pathway Tools software (<http://bioinformatics.ai.sri.com/ptools/>) that can be used for gap list generation and gap-filling for BioCyc databases [343]. Alongside these predictive algorithms, integrated omics data, as well as tissue-specific information can aid the process of gap filling by suggesting candidate metabolic reactions [344-346]. Finally, automatically generated gap-filling hypotheses should be validated by manual literature review and experimental data.

Despite the number of available tools and algorithms, there is no single method for complete gap filling of a given model. More complete genome annotations, metabolic databases as well as more powerful tools are currently required for elucidation of comprehensive metabolic models. As a result, there is currently no GEM that is complete and fully realistic.

#### **1.4.5 Towards global metabolic model of *Drosophila melanogaster***

Generation of GEMs of model organisms, including *Drosophila melanogaster*, has been recognised as a priority in systems biology [347]. It would improve our understanding of holistic function of this species, and as a result of all organisms, including humans. Identification of all metabolites and mapping them to metabolic pathways would allow a greater understanding of global metabolism of all species. GEM of the fruit fly would have numerous applications. First of all, it would be the first global metabolic model of an insect species, and hence it could be used as a basis for the development of GEMs of other insect, including pests and disease vectors, such as mosquito. Insect pests kill over 20% of the world's crops and kill millions of people every year. There is a high demand for new insecticides and insect GEMs have a potential to identify new drug targets and provide a tool for computational modelling of drug action. Secondly, by elucidation the core metabolome of *Drosophila*, various fly models of human diseases could be modelled and tested, in particular metabolic disease, such as IEMs. Finally, GEMs provide a powerful tool for systems biology and functional genomics. Metabolomics has a potential to elucidate functions of novel genes, including those coding for orphan enzymes.

There is currently no authoritative global metabolic map for *Drosophila melanogaster* or any other insect (<http://bigg.ucsd.edu>) [304]. Only partial reconstructions of the fly metabolism have been previously reported, including a constraint-based model of ATP-producing pathways, during hypoxia and aging [176, 348, 349]. A basic computational model of core metabolism of *Drosophila* has been generated by Dow/Davies lab (Dr D. Erben, personal communication). The stoichiometric model was built using *D. melanogaster* BioCyc database version 4.0.1.1 and curated using literature and KEGG (<http://www.genome.jp/kegg/>) database. The model was under the steady-state assumption [305], which means that there is no metabolite accumulation and the sum of reaction fluxes synthesising a compound must be the same as the sum of reaction fluxes breaking it down. The model is constraint-free (no biophysical constraints on reaction and transport rates were applied) [335] because there are currently no data on reaction rates in the fly. PathwayTools was then used to visualise and query the map during curation and to generate the list of metabolic gaps. The ability to produce biomass precursors with no priority (all components equally important) was assessed using COBRA toolbox for Matlab. The model identified 2484 metabolites, 2759 reactions and 145 metabolic gaps. Further literature review revealed that not all of these gaps are relevant to the fruit fly. The remaining gaps were searched against the *Drosophila* genome using the Protein Basic Local Alignment Search Tool (pBLAST), identifying 26 candidate fly genes with at least partial sequence similarity to the missing enzymes. In order to select the gaps that were the easiest to study, genes enriched in single tissues (FlyAtlas.org) and, for which knockout fly lines were available in stock centres, were selected. These genes were then correlated to human disease, with focus on IEMs, and a single gap, *CG30016* gene, has been selected (more details on *CG30016* are presented in Chapter 3). *Drosophila CG30016* gene was chosen because it has a high sequence homology with 5-hydroxyisourate hydrolase (5-HIU) of other species, which is involved to urate degradation pathway. Due to the absence of this pathway in humans high urate levels can lead to an IEM, hyperuricemia. Understanding the molecular basis of urate degradation in the fly could increase our knowledge on hyperuricemia and potentially lead to the development of *Drosophila* model of the disease. This could contribute to the development of much needed novel treatments for hyperuricemia. Moreover, *Drosophila CG30016* gene is highly enriched in the tubules, and *CG30016* knockout flies are available in stock centers.

## 1.5 Project aims

The aim of this project was to develop an efficient method to assess and fill in gaps of the primary global metabolic model of *Drosophila melanogaster* using metabolomics approaches combined with standard reverse genetics tools. Selected gap was 5-hydroxyisourate hydrolase (5-HIUH), which plays a role in urate degradation pathway. Inability to degrade urate can lead to IEMs in humans, including hyperuricemia. Filling in gaps of the fly metabolic map, would aid better understanding of urate degradation. This in turn could allow generation of the fly model of hyperuricemia, and development of new much needed therapeutics for the disease. Moreover, developing a method for filling in metabolic gaps, would allow fast and efficient gap-filling process. Our approach was to obtain a fly line, in which the expression of *5-HIUH* gene was disrupted, followed by its validation as a knockout. The next step was to assess the tissue, in which the gene was enriched in order to identify phenotypes. Finally, LC-MS metabolomics analysis was performed in order to determine whether the expected reaction was disrupted, and to observe the results of the knockout on the whole metabolism.

Moreover, we aimed to build a primary tissue-specific metabolome of the fruit fly, in order to establish whether different *Drosophila* tissues have different metabolomes and if they correspond to tissue-specific transcriptome of the fruit fly (FlyAtlas.org). This was done by dissecting different fly tissues and elucidating their complete metabolome using LC-MS metabolomics analysis.

Development of an efficient gap-filling protocol, together with tissue-specific metabolome of the fly, would aid the reconstruction of a more comprehensive *Drosophila* metabolic map. This in turn would allow testing the validity of a fruit fly as an *in vivo* model for human metabolism and metabolic imbalances including IEMs. It would also identify any limitations of this model. Understanding of the core of *Drosophila* metabolism is essential for this undertaking. Development of such powerful animal model would enable identification and testing of new drugs for IEMs and other metabolic imbalances in humans and other animals. Moreover, it would provide a holistic understanding of organismal function and increase understanding of the interplay between individual tissues in a whole organism.

## 2 Materials and Methods

All materials were obtained from Sigma (Sigma Aldrich, UK) unless otherwise specified.

### 2.1 *Drosophila* maintenance and stocks

#### 2.1.1 Fly stocks

Fly lines used in the following experiments were either obtained from publicly available stock centres, or generated in-house. All fly lines are described in the table below (Table 2-1), including their fly ID, genotype, description and the source.




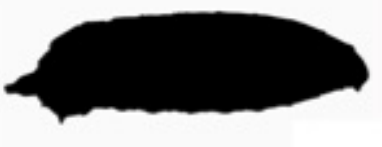
#### 2.1.2 Fly maintenance

*Drosophila melanogaster* flies were maintained in standard conditions unless otherwise specified. Flies were reared on standard yeast cornmeal, sucrose and agar medium (Appendix I for full medium recipe) in ventilated plastic vials. The insectary conditions were tightly controlled, with temperature of 23°C, 55% atmospheric humidity and a 12:12 h light:dark cycle. Adult flies were transferred to fresh vials every two weeks unless flies of specific age were required. In order to collect adult flies of specific age, a laying population of around 20 males and 20 females were transferred to fresh vials every day. The progeny was then collected in fresh vials when adults emerged (1 day old). This allowed precise age determination of adult flies. For most experiments, flies of 5-7 days were used. Where larvae were used, they were collected directly from the medium, on which egg laying flies were feeding. Larval stages and corresponding sizes are represented in Table 2-2.

**Table 2-1 Fly lines used in this study. Each fly line used is described including its ID, genotype, brief description and source. Flies were either made in the lab or purchased from stock centres.**

Fly ID	Genotype	Description	Source
Canton S	<i>w<sup>+</sup>; +/+; +/+</i>	Wild type – <i>Drosophila melanogaster</i>	Bloomington Drosophila Stock Center
BDSC 17767	<i>w<sup>1118</sup>; PBac[279]<i>hbs</i><sup>c06523</sup></i>	PiggyBac insertion in <i>w<sup>1118</sup></i> background targeting <i>hbs</i> gene. No effect on <i>hbs</i> gene expression	Bloomington Drosophila Stock Center
BDSC 18554	<i>w<sup>1118</sup>; PBac{WH}<i>CG30016</i><sup>f02466</sup></i>	PiggyBac insertion in <i>w<sup>1118</sup></i> background targeting <i>CG30016</i> gene. Loss of function mutation	Bloomington Drosophila Stock Center
BDSC 18814	<i>w<sup>1118</sup>; PBac{WH}<i>Uro</i><sup>f04888</sup></i>	PiggyBac insertion in <i>w<sup>1118</sup></i> background targeting <i>Uro</i> ( <i>urate oxidase</i> ) gene. Loss of function mutation	Bloomington Drosophila Stock Center
BDSC 225	<i>Ry<sup>506</sup>(CS)</i>	Deletion of ~1/3 coding region in <i>ry</i> ( <i>rosy, xanthine dehydrogenase</i> ) in a Canton S background. Loss of function/amorphic allele.	Bloomington Drosophila Stock Center
c42-GAL4	<i>w<sup>-</sup>; +/+; c42-GAL4/c42-GAL4</i>	GAL4 enhancer trap specific to the tubule principal cells	(Sozen et al., 1997); Dow/Davies Labs
capaR-GAL4	<i>w<sup>-</sup>; +/+; c42-GAL4/capaR-GAL4</i>	GAL4 enhancer trap specific to the tubule cells	Selim Terhzaz, Dow/Davies Lab
12888-3-8M-Ch2	<i>w<sup>-</sup>; UAS-<i>CG30016.His6</i>/Cyo</i>	<i>CG30016</i> tagged with 6xHis under UAS control	Dominika Korzekwa, Dow/Davies Lab
12888-3-10M-Ch3	<i>w<sup>-</sup>; +/+; UAS-<i>CG30016.His6</i>/TM3 Sb</i>	<i>CG30016</i> tagged with 6xHis under UAS control	Dominika Korzekwa, Dow/Davies Lab
UAS-RelHis-C42Gal4	<i>w<sup>*</sup>; {UAS-<i>Rel.His6</i>}<sup>2</sup>/Cyo; GAL4<sup>c42</sup>/TM3 Sb</i>	<i>Relish</i> tagged with 6xHis under UAS control driven by c42 GAL4	Dow/Davies Lab
UAS-pHluorin	<i>w<sup>1118</sup>; +/+; UAS-pHluorin(Perox)/TM3</i>	pHluorin fusion with upstream UAS; localizes to peroxisomes	Dow/Davies Lab

**Table 2-2 Developmental *Drosophila* larval stages. Different stages of fly larvae were used in the study; larva size and picture is illustrated for each stage.**

	Larval stage	Size (length in mm)
	First instar (L1)	~1.0
	Second instar (L2)	~2.0
	Early third instar (L3E)	~3.0
	Third instar (L3)	~3.8

### 2.1.3 Fly mating

After constructing appropriate transgenic plasmids as described in Section 2.8.6, UAS/GAL4 system (Section 1.1.4) was employed to express transgenes in a desired tissue. Female virgin flies of UAS-transgene line were crossed to GAL4 males. In order to collect adult virgin flies, a laying population of around 20 males and 20 females were transferred to fresh vials every day. The progeny was then collected in fresh vials when adults emerged early in the morning to select female virgin flies. 5-10 virgin flies of UAS-transgene strain were then mated to 10-20 GAL4 male flies. They were transferred to a new vial every 2-3 days and tested for the presence of the transgene by visual markers and qPCR (quantitative PCR).

## 2.2 *Drosophila* tissue dissection

Seven-day old were anaesthetised briefly on ice and immediately used for tissue dissection. Tissues were dissected in *Drosophila* Schneider's medium (Invitrogen, UK). Tissues were either directly transferred to slides containing a drop of PBS (Phosphate-buffered saline pH 7.4: 137 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>) for Immunocytochemistry/microscopic visualisation or transferred every 30 min to appropriate buffers for other experiments. Male and female tissues were collected separately.



## **2.3 RNA extraction**

RNA extraction was performed in a nuclease-free environment. Nuclease-free environment was achieved by working and storing all equipment used for RNA extraction, in a specifically designated RNase (ribonuclease)-free zone of the lab. All working surfaces were washed using RNaseZap® reagent (Thermo Fisher Scientific, UK) before and after each experiment. Gloves were used at all times and replaced frequently. Tubes and tips used in the experiments were certified RNase-free and came from an unopened sterile box. The tips were RNase-free barrier pipette tips to ensure no cross-contamination of RNA samples. Water and buffers were Diethylpyrocabonate (DEPC) – treated in order to inactivate RNases.

### **2.3.1 RNA extraction using whole flies and whole larvae**

Seven-day old adult flies (five males and five females) were collected and anaesthetised under CO<sub>2</sub>. They were then homogenised using a micropestle in 250 µl Trizol (Life Technologies, UK), followed by 10 second of sonication using ultrasonic cell disruptor (Misonix, Inc., USA). Additional Trizol was added to a total volume of 1 ml following incubation at room temperature (RTm) for 5 min and addition of 200 µl of Chloroform. The homogenates were then vortexed for 15 sec and subsequently centrifuged at 12000 g and 4°C for 15 min. The aqueous upper layer was then transferred to a new 1.5 ml microcentrifuge tube without disturbing the interphase. Isopropyl alcohol (½ of total sample volume) was then added followed by precipitation at RTm for 10 min. The samples were then centrifuged at 12000 g and 4°C for 10 min. After removing the supernatant, the pellets were washed with 70% ice-cold ethanol and vortexed. Following centrifugation at 8000 g and 4°C for 5 min, the supernatant containing ethanol was discarded without disturbing the pellet. In order to remove all ethanol, the pellets were air dried at RTm for ~5 min. They were then resuspended in 30 µl of RNase-free water and gently mixed. Extracted RNA was immediately used for cDNA synthesis. Remaining RNA was stored at -80°C.

Where larvae were used, 10 third instar feeding larvae were collected. The same RNA extraction protocol was followed as with adult flies.

### **2.3.2 RNA extraction using Malpighian tubules**

Malpighian tubules were dissected as described in paragraph 2.2.2. Seven-day old flies (30 males and 30 females) were collected and used for dissection. Dissected tubules were homogenised using a micropestle in 250 µl RLT lysis buffer (Qiagen, UK) containing 10% β-mercaptoethanol, followed by 10 second of sonication using ultrasonic cell disruptor (Misonix, Inc., USA). The homogenates were then centrifuged for 5 min at 4°C and 13000 g and the supernatants were collected and transferred to a new 1.5 ml microcentrifuge tube. RNA was extracted using the Qiagen RNeasy Mini Extraction Kit (Qiagen, UK) following manufacturer's instructions. DNA digestion (Qiagen RNase-Free DNase Set) was performed in order to reduce genomic DNA contamination. RNA was eluted from the column using 30 µl of nuclease-free water and immediately used for cDNA synthesis. Remaining RNA was stored at -80°C.

### **2.3.3 Complementary DNA synthesis**

Complementary DNA was synthesised using 500-1000 ng of RNA using Superscript II reverse transcriptase (SuperScript® II, Invitrogen, UK). The synthesis of cDNA was performed following the manufacturer's protocol and the total reaction volume was 20 µl. Synthesised cDNA was stored at -20°C until further use. The cDNA was used for standard PCR and qPCR described in sections below.

In order to control for genomic contamination, samples were prepared without SuperScript II. In these cases, cDNA was not synthesised. Hence, any product detection in PCR and qPCR experiments was due to genomic contamination rather than amplification of cDNA.

### **2.3.4 Quantification of nucleic acid**

Both RNA and cDNA were quantified using the NanoDrop™ 1000 spectrophotometer (Thermo, UK) in 2 µl sample volume. NanoDrop analyser uses 1.0 mm and 0.2 mm path lengths, which makes it more robust than a standard spectrophotometer with a path length of 10.0 mm. Background absorbance was measured using a reference sample (water sample or the sample elution buffer) in order to provide the 'blank' reading. The absorbance was measured at two wavelengths:  $A_{260}$  and  $A_{280}$  nm. Sample concentration was calculated using the equation below:

$$C = (A * e)/b$$

where C corresponds to the nucleic acid concentration in ng/μl, A corresponds to the absorbance in AU, e is the extinction coefficient in ng-cm/μl (which is 50, 33 and 40 for double-stranded DNA, single-stranded DNA and RNA respectively), and b corresponds to the path length in cm.

Sample purity was also monitored using the  $A_{260}/A_{280}$  ratio. For RNA ratio values of ~2.0 and for DNA values of ~1.8 were considered pure. In cases where ratio values were considerably lower or higher, the samples were discarded.

## 2.4 Oligonucleotide synthesis

Oligonucleotides were used as PCR and qPCR primers. The primers were designed and tested using the software MacVector 11.1.1 (MacVector, Inc., UK). They were then synthesised using Integrated DNA Technologies® (IDT) services. Provided oligonucleotides were desalted and scaled up to 1 μmol by IDT, and the quality was monitored by mass spectrometry. Purchased primers were resuspended in IDTE buffer (10 mM Tris, pH 7.5, 0.1 mM EDTA) to obtain stock concentration of 100 μM. Primers were stored at -20°C at a working concentration of 10 μM.

## 2.5 Polymerase chain reaction (PCR)

### 2.5.1 Standard PCR

Standard PCR was performed using DreamTaq Green PCR Master Mix (Thermo, UK). The master mix contained DreamTaq DNA Polymerase, optimised DreamTaq buffer (two tracking dyes and density reagent), 4 mM MgCl<sub>2</sub>, and dNTPs (0.4 mM each). Each reaction contained the master mix, forward and reverse primers, template DNA and nuclease-free water and the total reaction volume was 25 μl. The reactions were set up according to the manufacturer's recommendations. Thermal cycling conditions are described in the table below (Table 2-3).

**Table 2-3 Standard PCR thermal cycling conditions.**

Step	Number of cycles	Temperature	Time
Initial denaturation	1	95°C	5 min
Denaturation	30	95°C	30 sec
Annealing		55-65°C (T <sub>m</sub> -5)	30 sec
Extension		72°C	1 min/kb
Final extension	1	72°C	10 min

DreamTaq DNA Polymerase is an enhanced high-efficiency Taq DNA Polymerase and has 5' to 3' polymerisation and exonuclease activity and lacks proofreading activity. PCR products were analysed using agarose gel electrophoresis (described below).

Gradient PCR using DreamTaq enzyme was performed to test PCR and qPCR primers and establish their annealing temperature. Once the annealing temperature was confirmed and a single band was isolated on an agarose gel, Herculase II Fusion DNA Polymerase (Agilent Technologies, UK) was used to amplify the product of interest. The Herculase II is a *Pfu* DNA Polymerase based enzyme. It has a significantly higher fidelity than Taq Polymerase, results in higher yields and can be used for longer PCR products. Hence, it was the preferred enzyme for cloning experiments. The reactions were set up according to the manufacturer's protocol.

### **2.5.2 Quantitative reverse-transcriptase PCR (qPCR)**

Quantitative PCR is a powerful technique used to determine gene expression of a gene of interest. QPCR was performed using TaqMan® Gene Expression Master Mix and Applied Biosystems StepOne™ Real-Time PCR System. Reaction mix contained AmpliTaq Gold® DNA Polymerase, UP (Ultra Pure), Uracil-DNA Glycosylase (UDG), deoxyribonucleotide triphosphates (dNTPs) with deoxyuridine triphosphate (dUTP), ROX™ Passive Reference, and optimised buffer. Components of each reaction are represented in a table below (Table 2-4).

**Table 2-4 QPCR reaction components.**

Reaction component	Quantity
TaqMan® Gene Expression Master Mix	5 µl
cDNA template	2 µl
Primers and probe	0.5 µl
Water	2.5 µl
<b>Total reaction volume</b>	<b>10 µl</b>

PCR products were quantitatively synthesised from cDNA templates using AmpliTaq DNA Polymerase. The enzyme is suitable for Hot Start PCR and is a form of Amplu Taq® DNA Polymerase. This means that it is only active at temperatures where DNA templates are fully denatured. It is also more pure than Amplu Taq® DNA Polymerase, which prevents bacterial contamination and inaccurate quantification of PCR products.

TaqMan primers and probes were designed and generated as described in Section 2.4 and their sequences are provided in Appendix II. They contained a 6FAM™ reporter dye at the 5' end and a nonfluorescent quencher at the 3' end. The increase in fluorescence of the reporter dye allows the quantification of PCR products. QPCR cycling parameters used here were consistent with the manufacturer's protocol.

Several types of controls were used. In order to control for genomic contamination negative controls prepared during cDNA synthesis without the Superscript II were used as a template. Moreover, reactions without any template were prepared in order to monitor and subtract background fluorescence from the experimental samples. Endogenous controls were prepared for each sample to normalise the data and account for variability in the initial concentration and quality of cDNA. Reference control samples contained  $\alpha$ -Tubulin84B primers.  $\alpha$ -Tubulin84B gene is constitutively expressed. Four replicates were prepared for each reaction.

### 2.5.3 Data analysis

QPCR results were analysed using the StepOne™ Software v2.1. The analysis was based on the  $2^{-\Delta\Delta CT}$  method [350] and relative gene expression levels (either percentage expression or fold changes) were calculated. Statistical analyses, including P-values and

standard error of the mean (SEM), were performed using GraphPad Prism software (GraphPad Software, USA).

## **2.6 Agarose gel electrophoresis**

PCR products were separated on a 1% agarose gel in order to determine their specificity and quality. Agarose gels were prepared in 0.5x TBE buffer containing 90 mM Tris, 90 mM boric acid (pH 8.3), 2 mM EDTA and 0.1 µg/ ml EtBr (ethidium bromide). Where required, 5x loading dye was added. 20-30 µl samples and 5 µl 1 kb ladder (Invitrogen, UK) were loaded into the wells.

Gels were run at 100 V and they were visualised using high performance ultraviolet transilluminator (UVP, UK). PCR product band sizes were compared to the ladder. If required, PCR product bands were extracted from the gel as described in the section below.

## **2.7 PCR/Gel purification**

PCR products excised from gels were purified using QIAquick Gel Extraction Kit (Qiagen, UK) following manufacturer's protocol. PCR products were purified using QIAquick PCR Purification Kit (Qiagen, UK) following manufacturer's protocol. DNA was eluted in 30 µl of nuclease-free water and quantified using NanoDrop as described in Section 2.3.4.

## **2.8 Molecular cloning**

Molecular cloning protocols were used in order to clone *Drosophila melanogaster* genes. They were then expressed using several different expression systems described below. Expressed proteins were quantified using Bradford Assay and visualised using Western blotting and immunocytochemistry (ICC). Where required, proteins were purified and used in enzymatic assays.

### 2.8.1 Bacterial strains and plasmids

Table 2-5 Bacterial strains and vectors used in the study.

Bacterial strain	Use
One Shot® TOP10 chemically competent <i>E. coli</i> (Invitrogen)	Stable propagation and maintenance of recombinant plasmids
BL21 Star™(DE3)	Regulated expression of recombinant plasmids
Plasmid	Use
pF25A ICE T7 Flexi® Vector	Cell-free expression system
pET101/D-TOPO®	Expression in <i>E. coli</i>
pMT/V5-His-TOPO® vector	Expression in S2 cells
pUAST	UAS/GAL4 binary induction of transgenes <i>in vivo</i>

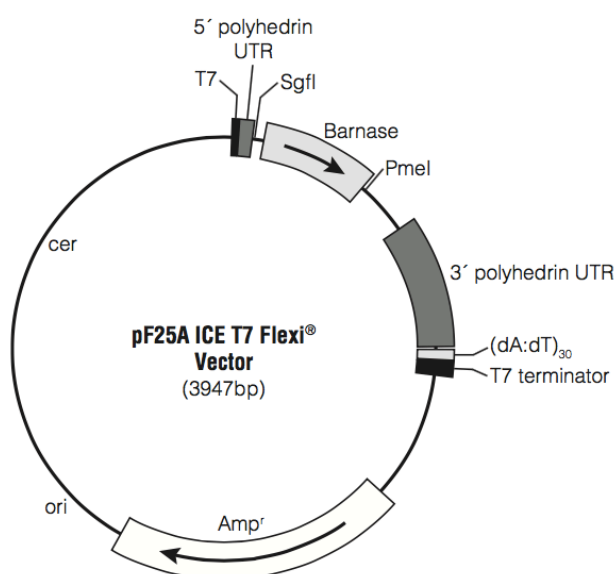
### 2.8.2 Construct generation

Recombinant plasmid constructs were generated for *Drosophila* gene expression using several expression systems described below. Standard cloning procedures were performed. *Drosophila* genes were amplified using Herculase II DNA Polymerase as described in Section 2.5. PCR products were then separated on an agarose gel and gel purified. They were then digested using specific restriction enzymes (purchased from New England Biolabs or Promega) following manufacturer's protocol and PCR purified. Plasmids were also digested with the same restriction enzymes, CIP (Calf Intestine Phosphatase) treated (NEB, UK) and gel purified. Ligation reactions were performed using Quick Ligation™ Kit (NEB, UK) and following manufacturer's protocol. Several molar ratios of PCR product:vector were used for optimal results. Generated constructs were then used for transformation of competent cells described in Section 2.8.7.

### 2.8.3 Cell-free protein expression system

TNT® T7 Insect Cell Extract Protein Expression System (Promega, UK) is a quick and simple expression system for cell-free expression of proteins. It employs TNT technology and a coupled transcription/translation reaction, which makes it faster than other expression systems. The system was used for *Drosophila melanogaster* protein expression followed by protein purification (described in Section 2.10). Vector used in the experiment was pF25A ICE T7 Flexi® Vector illustrated in Fig. 2.1 (Promega, UK) and all steps were

performed in ribonuclease-free environment following manufacturer's protocol.



**Fig 2.1 Map of pF25A ICE T7 Flexi® Vector.** The vector was used for *Drosophila* gene cloning and expression using TNT® T7 cell-free expression system. It contains SgfI and PmeI restriction sites and ampicillin resistance gene.

<https://www.promega.co.uk/resources/protocols/product-information-sheets/g/pf25a-ice-flexi-vector-protocol/>

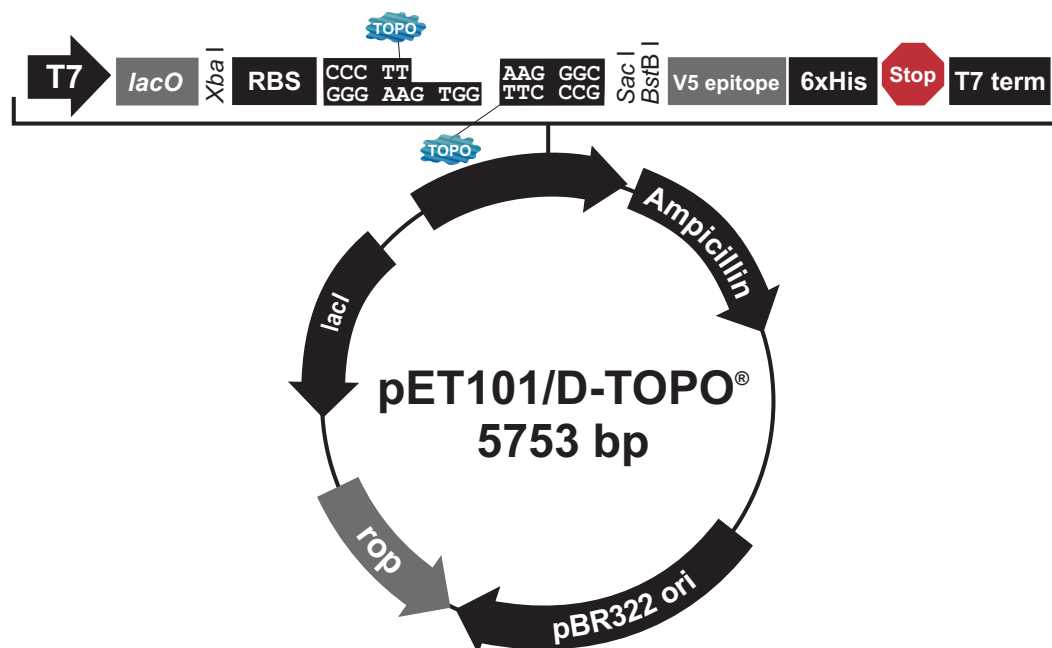
#### 2.8.3.1 Cloning primers and PCR amplification

Oligonucleotide primers were designed as described in Section 2.4. Forward primer contained SgfI restriction site, and the reverse primer contained PmeI restriction site as well as 6xHis tag for detection. Complete primer sequences are available in Appendix II. PCR amplification was performed using Herculase II DNA Polymerase as described in section 2.5.

#### 2.8.4 TOPO pET100 expression system

Champion™ pET100 Directional TOPO® Expression Kit (Invitrogen, UK) uses T7 RNA Polymerase and was used to express *Drosophila* genes in chemically competent *E. coli* cells, BL21 Star™(DE3) (Invitrogen, UK). The system allows regulation of T7 RNA polymerase expression. BL21 cells contain *lac* repressor, which prevents expression of T7 RNA polymerase. In order for the gene to be expressed, isopropyl β-D-thiogalactoside (IPTG) must be added. Vector used in the experiment was pET101/D-TOPO® (Figure 2-2). All reactions were performed according to manufacturer's protocol.





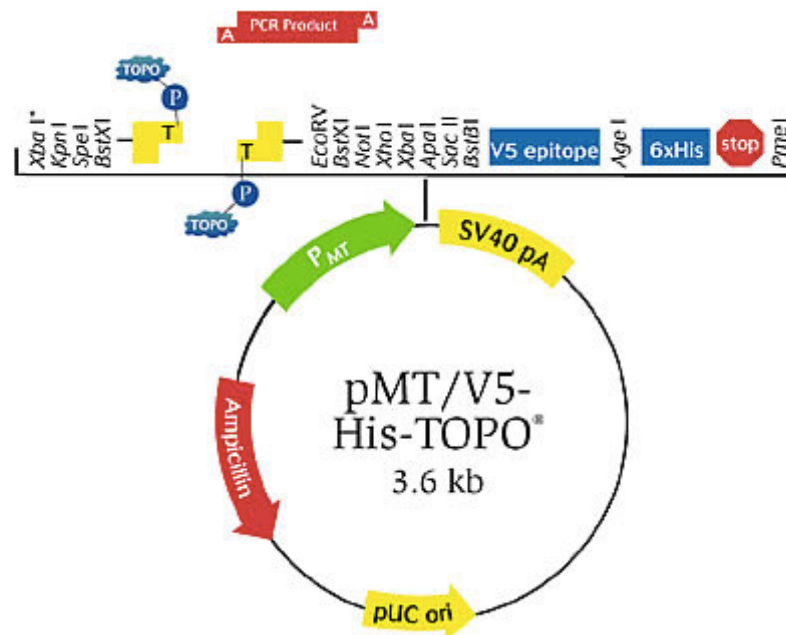
**Fig 2.2 Map of pET101/D-TOPO<sup>®</sup> plasmid.** The vector was used for *Drosophila* gene cloning and expression in *E.coli* cells. It contains multiple restriction sites, V5 epitope, 6xHis tag and ampicillin resistance gene.  
<https://www.thermofisher.com/order/catalog/product/K10101>

#### 2.8.4.1 Cloning primers and PCR amplification

Oligonucleotide primers were designed as described in Section 2.4. Forward primer was designed to contain CACC sequence required for directional cloning and EcoRI restriction site. Reverse primer contained XbaI restriction site and 6xHis tag for detection. Full primer sequences are available in Appendix II. PCR amplification was performed using Herculase II DNA Polymerase as described in section 2.5.

#### 2.8.5 DES TOPO cloning

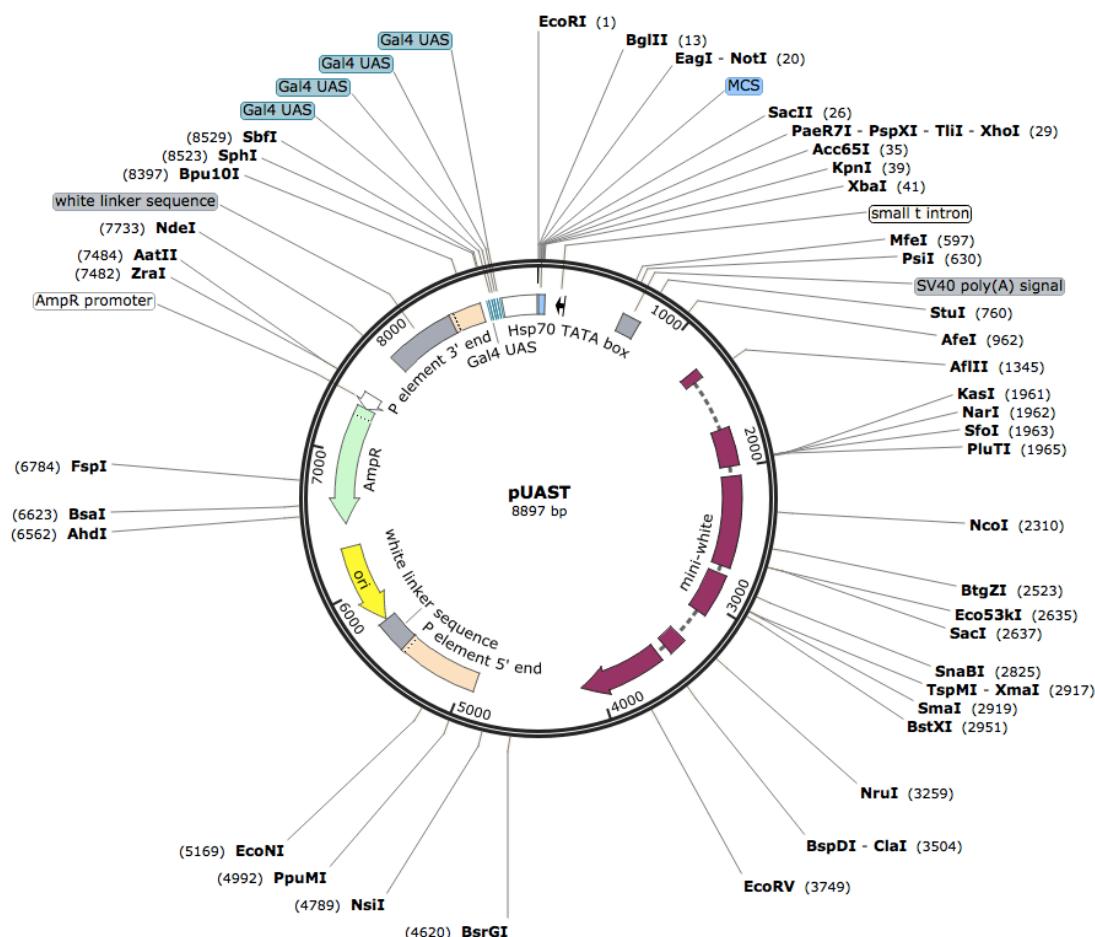
DES<sup>®</sup> TOPO<sup>®</sup> TA Expression Kit was used to generate constructs for transient inducible expression of *Drosophila* protein *in vitro* in S2 cells. This technology does not require the ligation step or post-PCR procedures. *Drosophila* genes were PCR amplified using Taq DNA Polymerase resulting in PCR products with polyA (Adenine) overhangs. PCR products were then directly cloned into the pMT/V5-His-TOPO<sup>®</sup> vector (see Figure 2-3). In order to induce protein expression in S2 cells, CuSO<sub>4</sub> was added to the cells to activate the metallothionein promoter. These constructs were used to determine protein subcellular localization as well as protein purification followed by enzyme assay.



**Fig 2.3 Map of pMT/V5-His-TOPO® vector.** The vector was used for gene expression in S2 cells. *Drosophila* genes were PCR amplified to contain polyA overhangs, which allowed direct cloning into Topoisomerase (TOPO) containing vector. TOPO catalyses bond formation between vector thymidines and polyA overhangs. The vector contains multiple restriction sites, V5 epitope, 6xHis tag and ampicillin resistance gene.  
<https://www.thermofisher.com/order/catalog/product/K412501>

### 2.8.6 Cloning into pUAST plasmid

*Drosophila* genes were cloned into the pUAST vector for germ line transformation of *Drosophila* embryos. These were subsequently used for UAS/GAL4 system induction of transgene expression *in vivo* in flies. Cloning into the pUAST vector (Figure 2-4) was carried out using standard cloning procedures. In order to confirm the insertion of transgenes into the pUAST vector, constructs were transiently expressed in S2 cells. Constructs were then sequenced and used for microinjection into *Drosophila* embryos as well as transformant identification and balancing (Best Gene, USA).



**Fig 2.4 Map of pUAST vector.** The vector was used for gene expression in S2 cells followed by microinjection into *Drosophila* embryos and *in vivo* transgene expression in the flies. The vector contains multiple restriction sites and ampicillin resistance gene.  
[http://www.snapgene.com/resources/plasmid\\_files/basic\\_cloning\\_vectors/pUAST/](http://www.snapgene.com/resources/plasmid_files/basic_cloning_vectors/pUAST/)

### 2.8.6.1 Cloning primers and PCR amplification

Oligonucleotide primers were designed as described in Section 2.4. Forward primer was designed to contain EcoRI restriction site. Reverse primer contained XbaI restriction site and 6xHis tag for detection. Full primer sequences are available in Appendix II. PCR amplification was performed using Herculase II DNA Polymerase as described in section 2.5.

### 2.8.7 Transformation of *E.coli* cells

TOP10 competent cells were transformed with generated constructs following manufacturer's protocol. 5 µl of ligation reaction was used to transform the cells. Cells were transformed with uncut plasmids and cut unligated plasmids to provide positive and

negative controls respectively. Transformation reactions (100 µl and 200 µl separately) and controls (30 µl) were spread on L-agar plates containing 100 µg/ml ampicillin or another appropriate antibiotic depending on the resistant marker of the plasmid, and incubated overnight at 37°C. Positive colonies were identified using antibiotic resistance markers.

For high yield protein expression of confirmed transgenic plasmids, BL21 cells were transformed with 5-10 ng of plasmid and incubated on ice for 30 min. Cells were then heat-shocked for 30 sec at 42°C and immediately transferred to ice. Following the addition of 250 µl of S.O.C medium at RTm to each samples, cells were incubated at 37 °C for 30 min with shaking (200 rpm). The entire transformation reaction was then added to 10 ml of LB containing the appropriate antibiotic and incubated overnight at 37°C with shaking.

### **2.8.8 Purification of plasmid**

Positive colonies were picked from agar plates and resuspended in 10 µl of water. They were then added to 5 ml of lysogeny broth (LB) containing appropriate antibiotic and incubated with shaking overnight at 37°C. Liquid colonies were then centrifuged to form a pellet and resuspended in DNA lysis buffer. Plasmid DNA was purified using QIAprep Spin Miniprep or Maxiprep Kit (Qiagen, UK) following manufacturer's protocol. Plasmid DNA was eluted using 30 µl (Miniprep) or 500 µl (Maxiprep) of water or TE buffer.

### **2.8.9 Validation of cloning products**

In order to determine whether intact PCR products were present in the construct in the right orientation, validation experiments were carried out, including PCR, restriction digest and sequencing.

#### **2.8.9.1 PCR**

PCR was performed in order to establish whether full-length transgenes were inserted in the correct orientation. Oligonucleotide primers were used: one primer starting in the transgene and the other one in the cloning vector. The results were monitored using gel electrophoresis.

### **2.8.9.2 Restriction digest**

Restriction digests were performed in order to determine the orientation of inserts. Restriction products were monitored on an agarose gel. Different size bands were observed depending on the orientation of the transgene.

### **2.8.9.3 Sequencing**

In order to determine whether the full-length transgenes were inserted in the vector and no mutations occurred during PCR, DNA sequencing was performed. Sequencing primers were designed; both in the vector in order to obtain the full transgene sequence. Sequencing was carried out by GATC Biotech, UK. Obtained sequences were analysed using Mac Vector Software. Glycerol stocks (20% peptone, 40% glycerol) of correct constructs were stored at -80°C.

### **2.8.10 Pilot expression of *Drosophila* protein in BL21 cells**

After overnight incubation of transformed BL21 cells, 500 µl of cells were added to 10 ml of LB containing the appropriate antibiotic were grown at 37°C with shaking till they reached density of 0.6-0.8 at OD<sub>600</sub> (typically 1-2 h).

Once the right cell density was reached, 1 ml of each sample was collected and centrifuged at 8000 rpm for 5 min. Pellets were then frozen at -20°C and subsequently used in SDS-PAGE and Western blotting as a negative control (uninduced sample).

The remaining 9 ml of cells were used for pilot expression. In order to induce *Drosophila* protein expression, 1 M IPTG was added to a final concentration of 0.4 mM. The expression was performed at two different temperatures (25°C and 37°C) for 3 h or overnight in order to establish the best expression conditions. After the expression, OD<sub>600</sub> was measured and 1 ml of each sample was centrifuged at 8000 rpm for 5 min. Pellets were then stored at -20°C and subsequently used for SDS-PAGE and Western blotting.

Once the best expression conditions were determined, the transformation reaction of BL21 cells was scaled up to 500 ml. The protein expression was then induced using IPTG as described above and cells were grown in optimal expression conditions. Cells were then harvested by centrifugation at 3000 g for 10 min at 4°C and stored at -80°C. Pellets were resuspended in 20 ml of TBS containing 1 mg/ml protease inhibitor cocktail and lysed using French press at 8000 psi. Samples were then centrifuged at 13000 rpm for 30

min at 4°C. The supernatant was transferred to a fresh tube and centrifuged at 5000 rpm for 50 min. Pellets were discarded and the supernatant stored at -80°C till it was used for protein purification and enzyme assay.

## 2.9 *Drosophila* S2 Cell Techniques

### 2.9.1 Maintenance

*Drosophila* Schneider 2 (S2) cells (Invitrogen, UK) were maintained according to manufacturer's protocol. All procedures were carried out in sterile conditions. Cells were maintained in complete Schneider's medium (CSM) containing 10% fetal calf serum (FCS) at 25°C. Cell density was monitored daily until it reached  $10^7$  cells/ml. They were then passaged by diluting 6 ml of cells into 9 ml of CSM. Cells were maintained in 15 ml volumes in T75 flasks.

### 2.9.2 Transient transfection

S2 cells used for transient transfection were transferred to tissue culture six-well plates. When they reached the density of  $6 \times 10^6$  cells/ 3 ml, they were seeded into individual wells. 24 hours later transient-transfection was carried out using calcium phosphate transfection method following manufacturer's protocol. Plasmid DNA was prepared using a maxi-prep kit (Qiagen) and eluted in TE buffer. Transient transfection reaction is described in Table 2-6

**Table 2-6 Reaction components and the quantities used for transient transfection of S2 cells.**

Reaction component	Quantity
Plasmid DNA	19 µg
Plasmid DNA (GAL4 transgene, if required)	1 µg/ µl
CaCl <sub>2</sub>	240 mM
<b>Total reaction volume</b>	300 µl

Reactions were set up and well mixed followed by drop-wise addition to 300 µl of 2x HEPES buffered saline (HBS) (50 mM HEPES, 1.5 mM Na<sub>2</sub>HPO<sub>4</sub>, 280 mM NaCl, pH 7.1) and

mixing by bubbling air through with a sterile Pasteur pipette. The DNA/calcium phosphate was then incubated at RTm for 30 min to precipitate. It was then added drop-wise to the S2 cells with continuous gentle mixing by swirling. Cells were then incubated overnight at 25°C. Following the overnight incubation, transfected cells were mixed and resuspended followed by centrifugation at 1500 g at RTm for 1 min to form a pellet. Pellets were then resuspended in 3 ml fresh CSM. The centrifugation and resuspension steps were repeated three times. Following final resuspension in 3 ml CSM, cells were transferred into the six-well plates. In order to induce the expression of *Drosophila* transgenes, CuSO<sub>4</sub> (final concentration of 0.5 mM) was added to the cells. They were incubated at 25°C for 24-48 h to allow expression. Finally, cells were centrifuged at 1500 g for 1 min at RTm to harvest, washed in PBS and if not used immediately, stored at -80°C.

## **2.10 Protein purification and analysis**

Following *Drosophila* gene cloning and expression, proteins were purified and analysed as described below.

### **2.10.1 Protein purification**

Constructs were generated to contain 6xHis tags to allow protein purification and visualisation. Protein purification was performed using HisPur Ni-NTA Spin Columns (Pierce, UK) following manufacturer's protocol. His-tagged proteins were eluted in 100 µl of elution buffer. Protein concentrations were then measured using NanoDrop.

### **2.10.2 Changing protein buffer and concentration**

Where required, protein buffer was changed using Slide-A-Lyser® Dialysis Cassette (Thermo Scientific, UK) following manufacturer's protocol. Proteins were concentrated using VivaSpin500, 10 kDa MWCO (GE Healthcare, UK) following manufacturer's protocol.

### **2.10.3 Bradford assay**

Following protein purification, the protein concentration was quantified using the Bradford protein assay. The assay was performed in 96-well culture plates. Bovine serum albumin (BSA) protein standards were prepared of 0-5 µg BSA in water. These were set up in triplicates in a final volume of 50 µl. Typically, eight different standard concentrations

were used to establish the concentration of expressed *Drosophila* proteins. Several different dilutions of protein were prepared and 5 µl of each sample were set up in triplicates in a final volume of 50 µl. Subsequently, 200 µl Bradford reagent concentrate (BioRad, UK), diluted 1 in 5, was added to all the samples and standards. The absorbance at 590 nm was monitored using Mithras LB 940 Multimode Microplate Reader (Berthold Technologies). Absorbance readings of samples were then plotted against known concentrations of standards to interpolate sample absorbance and calculate their concentration.

#### 2.10.4 SDS-PAGE separation

In order to determine the size of expressed protein, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed. Electrophoretic separation was carried out using Novex NuPAGE electrophoresis system (Life Technologies, UK). 12-well 4-12% Bis- Tris-HCl (Bis(2-hydroxyethyl) imino-tris (hydroxymethyl) methane-HCl) buffered (pH 6.4) polyacrylamide gels were used. The running buffer was prepared by diluting 20x 20x NuPAGE™ MOPS SDS running buffer stock solution. Electrophoresis was performed at a constant voltage of 200 V for about an hour.

Samples were prepared as described in Table 2-7. Samples were then mixed and incubated at 70°C for 10 min followed by loading into gel wells. In order to determine the size of the protein, Novex® Sharp Pre-Stained Protein Standard (Life Technologies, UK) was also loaded into gel wells.

**Table 2-7 Sample components and quantity for SDS-PAGE.**

Sample component	Quantity
Protein	20 µg
NuPAGE LDS sample buffer (4x)	5 µl
<b>Total reaction volume</b>	20 µl

#### 2.10.5 Western blotting

SDS-PAGE separated proteins were transferred onto a Hybond ECL membrane (GE Life Sciences, UK). The membrane was activated by incubation in methanol for 2 min. The



transfer buffer used was Nupage Transfer Buffer (Life Technologies, UK) with 5% methanol. The transfer was performed at 100 V for 1 hour. Following the transfer, Ponceau S staining was performed for rapid detection of protein. The membrane was incubated in Ponceau S solution for 5 min following washes with 5% glacial acetic acid and water. Subsequently, the membrane was washed three times with TBST (TBS (500 mM NaCl, 20 mM Tris-HCl pH 7.5) containing 0.1% Tween 20), blocked at RTm for 1 hour in TBST containing 5% non-fat milk powder (Marvel) and incubated with primary anti-Histidine antibody overnight at 4°C (for details on used antibodies and their concentrations see Appendix III). Following the overnight incubation, the membrane was washed in TBST 3-5 times, blocked, incubated with a secondary antibody (Appendix III), washed three times in TBST and washed in TBS for 10 min. Chemiluminescence detection was performed using Amersham ECL Western Blotting Detection Reagents following manufacturer's protocol.

## **2.11 Immunocytochemistry**

Immunocytochemistry (ICC) staining was performed in S2 cells and dissected *Drosophila* tissues in order to determine the localisation of expressed protein *in vitro* and *in vivo*.

### **2.11.1 ICC of S2 cells**

Following transfection and protein expression in S2 cells, cells were mixed and resuspended and then collected in 15 ml falcon tubes. They were subsequently centrifuged at 3000 g for 5 min in a free rotating tabbletop centrifuge. Pellets were washed twice with PBS. 80 µl of cells were then transferred to Poly-L-lysine coated plates (Merk Millipore, UK) at a density of  $6 \times 10^6$  cells/ml and incubated at RTm for 30 min to adhere. Samples were fixed with 4% (w/v) paraformaldehyde in PBTA (PBS containing 0.5% Triton X-100) for 20 min and then washed three times with PBTA. Following 20 min blocking with PBTA containing 10% (w/v) goat serum, the cells were incubated with appropriate concentration of primary anti-His antibody (Appendix III) for 1 h at RTm. After three washes with PBTA, samples were blocked for another 30 min and incubated with appropriate secondary antibody for 1 h at RTm. Cells were then washed three times with PBTA, incubated with 100 µl of 500 ng/ml DAPI in PBS for 2 min. Following three washes with PBS, cover slips were mounted on slides using VectaShield medium (Vector

Laboratories, UK) and sealed with Glycerol Gelatin Finally, samples were visualised using confocal microscope (see Section 2.12 for details).

### **2.11.2 ICC of *Drosophila* Malpighian tubules**

Malpighian tubules were dissected in Schneider's medium and immediately transferred to Poly-L-lysine coated slides containing a drop of PBS (pH 7.4). Tubules were then fixed for 20 min with PBS containing 4% paraformaldehyde, washed three times with PBS and permeabilised for 30 min using 0.3% (v/v) Triton X-100 in PBS (PBT). Following 1 h blocking with PBT containing 5% (v/v) goat serum (PBT-GS), tubules were incubated overnight at 4°C with primary anti-His antibody at a desired concentration. Subsequently, samples were washed five times with PBT. Following 1-2 h blocking with PBT-GS, tubules were incubated overnight at 4°C with secondary antibody diluted to a desired concentration. The following day, samples were washed with PBT three times for 1 h per wash. They were then DAPI (4',6-diamidino-2-phenylindole)-stained (500 ng/ml DAPI in PBS) for 2 min and washed three times with PBS. Cover slips were then mounted on to the slides using VectaShield, sealed with glycerol-gelatine and visualised using confocal microscopy.

## **2.12 Imaging**

### **2.12.1 Fluorescent imaging**

Following immunocytochemistry, S2 cells or Malpighian tubules were imaged using confocal microscopy Zeiss 500 meta system (Zeiss, UK). In this study 40x objective was used. In order to image Alexafluor 568 antibody, HeNe1 543 nm laser was used, as well as 561-625 band pass filter. DAPI staining was visualised using UV source (mercury lamp) and the confocal photomultipliers. Images obtained from different channels were then merged using the LSM Meta Software.

### **2.12.2 Polarised light microscopy**

Polarising microscopy was used to visualise urate crystals. It is possible due to birefringent property of urate [351]. In order to visualise birefringent specimens, polarised microscope is equipped with two polarisers: one in the light path between the

light source and the specimen, and the other one (analyser) in the optical pathway between the objective and the observation tubes (Figure 2-5). The polariser can be rotated through 360 degrees, which enables regulation of the intensity of the polarisation light. When maximum brightness was achieved, samples were visualised. All samples were imaged using the same settings.

Using this technique, Malpighian tubules were assessed for the presence or absence of urate crystals. Tubules were dissected and slides were prepared as described in Section 2.2.

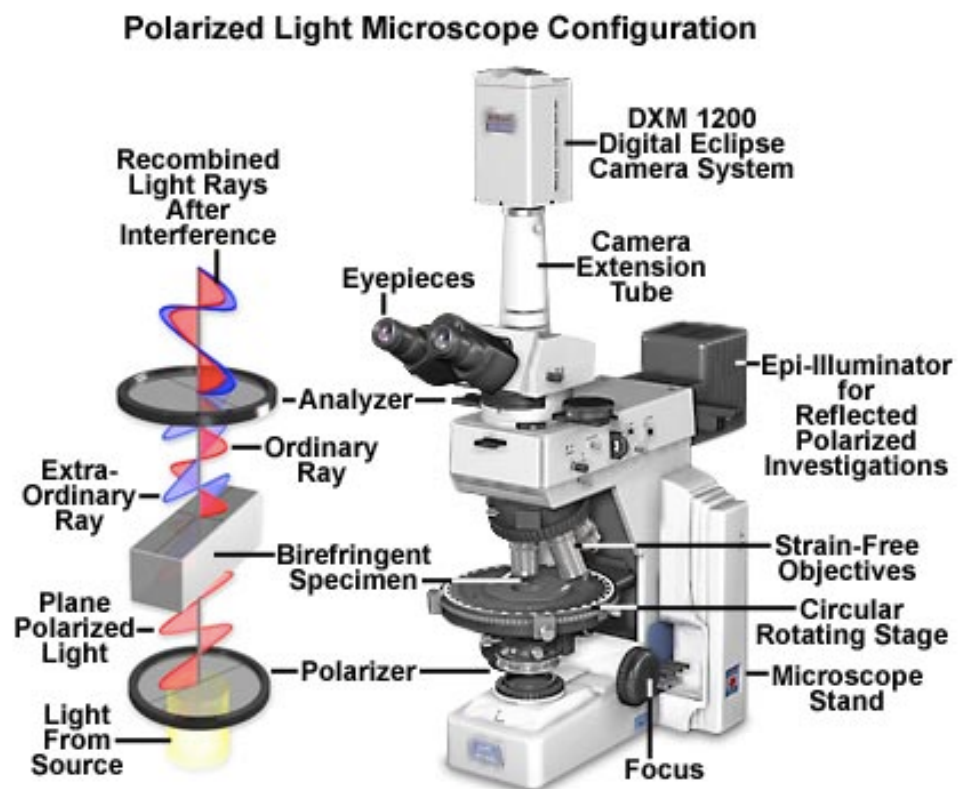


Fig.2-5 Polarised light microscope configuration

(<https://www.microscopyu.com/articles/polarized/polarizedintro.html>). Light passes through the polariser and the specimen before it hits the objective, the analyser and the observation tubes.

## 2.13 Stress and survival experiments

### 2.13.1 Oxidative stress survival assay

Five- to seven-day old flies were anaesthetised using CO<sub>2</sub> and females were separated from males. Thirty male and thirty female flies were collected in separate empty vials for 5-6 h to ensure that once placed in vials with survival food preparation, they eat immediately. Two different types of food containing different oxidising agents (either hydrogen peroxide or paraquat) were prepared as described in Table 2-8.

**Table 2-8 Types of food preparations used in oxidative stress survival assay. For each type, its components are described as well as**

Oxidising agent	Components	Mode of action
Hydrogen peroxide	1% UltraPure low melting point agarose (Thermo Fisher, UK)	Hydrogen peroxide is one of the simplest peroxides. Through catalysis it is converted to highly reactive hydroxyl radicals
	1% sucrose	
	1% hydrogen peroxide	
Paraquat	1% UltraPure low melting point agarose (Thermo Fisher, UK)	Paraquat is an electron acceptor and produces reactive oxygen species (ROS)
	1% sucrose	
	10 mM paraquat	

Flies were counted every 2-6 h and the number of dead flies was recorded. The numbers were then analysed using Kaplan-Meier survival curve as means  $\pm$  SEM (GraphPad Software). Median survival was also calculated using the same software. Appropriate controls were used and experiments were carried out in triplicates with n > 100.

### 2.13.2 Longevity assay

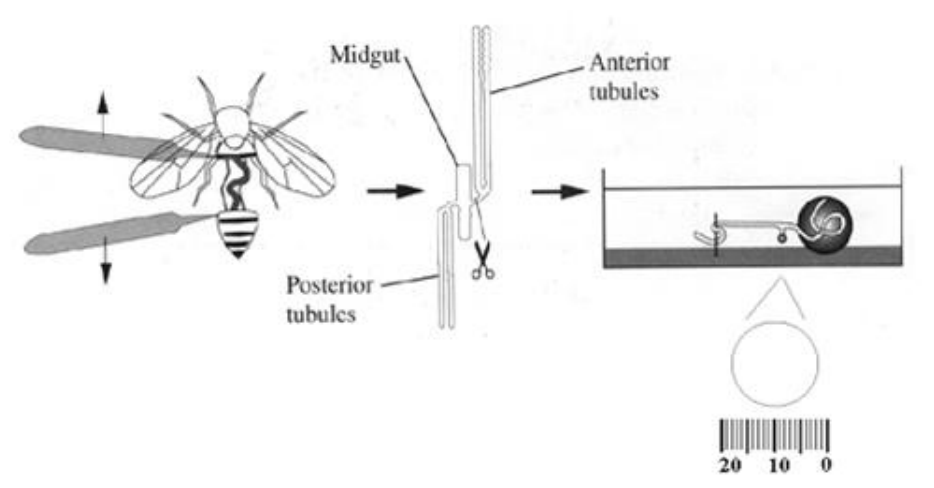
Flies used for the longevity assay were reared at 26°C. One-day old flies were collected from enclosing pupae over the period of 24 h. They were then mated for 48 h and separated into males and females. Ten male and ten female flies were collected in fresh food vials and kept at 26°C. They were transferred to fresh food vials every 2-3 days till all of them were dead. Number of dead flies was recorded every day. Data was analysed

using Kaplan-Meier survival curve as means  $\pm$  SEM (GraphPad Software) and median survival was calculated. Appropriate controls were used and the experiment was repeated three times with  $n > 100$ .

## 2.14 Fluid secretion assay

### 2.14.1 Principles of the assay

Fluid secretion assay (Figure 2-6) based on a classical Ramsay assay for tubule secretion was performed in order to establish secretion rates in different fly strains as described in Dow *et al.* [352]. Tubules were dissected with the ureter intact over a period of 30 min. One end of the tubule was wrapped around a metal pin and immersed in heavy mineral oil. The other end was placed in a small well filled with 10  $\mu$ l of *Drosophila* saline (Appendix IV): Schneider's medium (50:50) containing a small amount of Amaranth (red dye) to visualise secreted fluid. The whole tubule including the ureter was immersed in the oil. Pink bubbles emerging from the ureter were removed using a fine glass rod every 10 min over a period of 1 h and then measured using microscope eyepiece reticle. Data was analysed using Microsoft Excel (Microsoft Office Package).



**Fig. 2-6 Fluid secretion assay performed in *Drosophila*[352].** Malpighian tubules were dissected with intact ureter. One end of the tubule was then wrapped around a metal pin and the other end was immersed in *Drosophila* saline: Schneider's. Drops emerging from the ureter were measured every 10 min over a period of 1.5 h.

### **2.14.2 Collection of fluid for LC-MS**

For the metabolomics analysis using LC-MS, secretion assay was set up as described above. Drops emerging from the ureter were removed every 10 min and then pulled together and collected in 200 µl LC-MS solvent mixture (methanol: chloroform: water 3:1:1). Samples were stored at -80°C before LC-MS analysis was performed.

### **2.15 Haemolymph extraction**

Small 0.5 ml ependorf tubes were prepared and four holes were made in the bottom of each small tube using 0.8 x 40 mm needle (BD Microlance, UK). A drop of mineral oil was added to the bottom of 1.5 ml tubes. Small tubes were placed inside 1.5 ml tubes and placed on ice. Thirty flies were anaesthetised under CO<sub>2</sub>. They were then pierced with a 0.45 x 12 mm needle (BD Microlance, UK) between the thorax and the abdomen (below the wing) and placed in prepared tubes. Samples were centrifuged at 4°C and 5000 rpm for 1 min. This was repeated until 5 µl of haemolymph was obtained (around 4 times and 120 flies). The haemolymph collected in the mineral oil was then carefully pipetted using the P2 pipette and transferred to 200 µl LC-MS solvent mixture (methanol: chloroform: water 3:1:1). Samples were stored at -80°C before LC-MS analysis was performed.

### **2.16 Metabolomics analysis**

#### **2.16.1 Metabolite extraction**

Samples were collected of male and female fly tissues separately. Five whole flies or 60 tubules were collected in 1.5 ml microcentrifuge tubes containing 250 µl solvent mixture of methanol: chloroform: water in the ration of 3:1:1. They were then snap-frozen in liquid nitrogen, crushed with micropestles and sonicated for 20 sec using ultrasonic cell disruptor. Samples were centrifuged for 10 min at 10000 rpm at 4°C. The supernatant was transferred to a fresh tube and stored at -80°C. Four biological replicates were prepared. Quality control samples were prepared by mixing small aliquots of each sample. Blanks were also prepared by using solvent mixture on its own.

### 2.16.2 Liquid chromatography-Mass spectrometry (LC-MS)

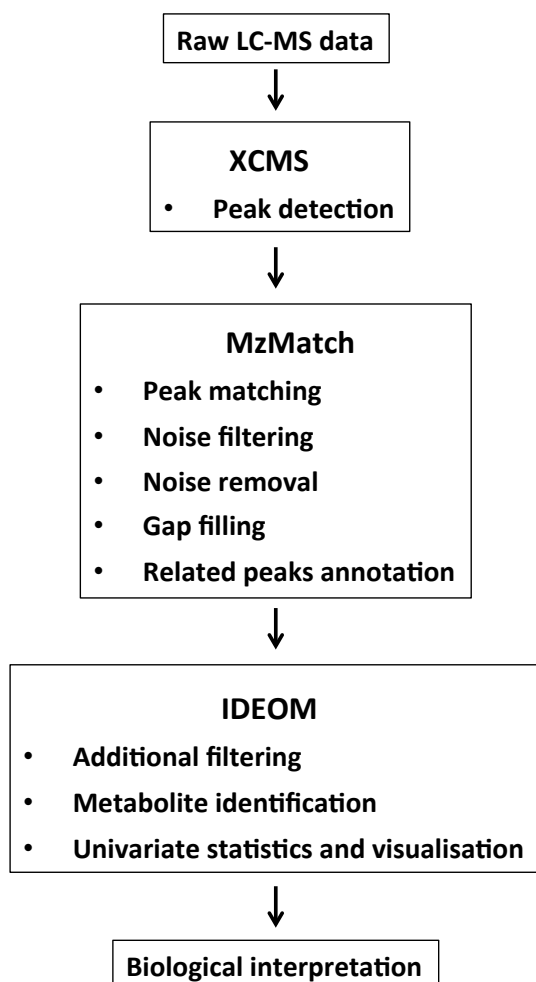
Samples were analysed at University of Glasgow Polyomics Facility using hydrophilic interaction liquid chromatography-mass spectrometry. The instrument used was Orbitrap™ Exactive™ mass spectrometer combined with UltiMate™ 3000 RSLC (rapid separation LC) separation system, which is a UHPLC system. The LC system was equipped with a 150 x 4.6mm ZIC-HILIC column running at 300 µl/min, and the separation was run in binary gradient mode. The mobile phase conditions were: solvent A 0.1% formic acid in HPLC grade water and solvent B 0.1% formic acid in acetonitrile; 80% B at (0 min) - 20% B at (30 min) – 20% B at (36 min) – 80% B at (37 min) – 80% B at (46 min). 10 µl of the sample was injected onto the column. The mass spectrometer was set at 50,000 resolution, and was run in positive and negative ionisation modes.

### 2.16.3 Data processing

Raw LC-MS data was processed using the IDEOM application (<http://mzmatch.sourceforge.net/ideom.php>) and default parameters [238]. The application consists of XCMS [228], MzMatch [239] and IDEOM processing [238] (Figure 2-7). XCMS and MzMatch were run in R. For peak detection, XCMS was used using default parameters, including PPM set at 2, and peak with range at min set at 2 s and max set at 100 s. MzMatch was used for peak matching (among the replicates), noise filtering and removal (peak shape, reproducibility, intensity, number of detections), gap filling, and related peaks annotation (adducts, isotopes, fragments etc.). Finally, IDEOM tools (in MS Excel) were used for additional filtering (samples allocation to experimental groups, contaminant filter applied to blanks, RSD as described in the Introduction), metabolite identification, and univariate statistics and data visualisation. Thanks to data export tools in IDEOM, data could then be exported for further analysis, including multivariate statistics (R algorithms, Metaboanalyst), pathway analysis (Metexplore, Pathos), etc. Statistical and pathway analyses performed for each dataset are described further in Chapter 5.2.

Metabolite identification was based on exact mass and RT<sub>m</sub>, where authentic standards were available. Otherwise, putative identification was based on exact mass and predicted RT<sub>m</sub>. Automated RT<sub>m</sub> prediction was performed as previously described in [353]. All masses were within 3 ppm of the exact molecular formula. This means that the only

competing metabolites within the database search were isomers. Metabolite levels for each experimental group were expressed as mean peak intensity (height) relative to the mean peak intensity of the control group. Metabolite levels were coloured according to relative intensity (blue = low, red = high).



**Fig. 2-7 IDEOM: data processing pipeline used for LC-MS data processing, adapted from [158].** Raw LC-MS data was uploaded to R followed by peak detection using XCMS, data processing steps in MzMatch and IDEOM, resulting in metabolite identification and data analysis using IDEOM tools as well as external tools.

Data obtained at Strathclyde University were analysed using Xcalibur software combined with Sieve Software 1.3 (Thermo Fisher Co.). Data were acquired and processed using Xcalibur software, which uses three user-created files: instrument methods, processing methods and sequences. Instrument methods are used to choose chromatographic and data acquisition settings for the LC and the MS respectively. Subsequently, processing methods are used to analyse the raw files. These include peak detection parameters, such as baseline settings and the void time (the elution time of an unretained peak) to



calculate the relative retention time of peaks. Finally, sequences specify volumes of injected samples, names of each sample and sample types, including QC samples, standards and blanks. Extracted ion chromatograms were analysed using Sieve software, an automated data mining and identification tool. Sieve performs statistical analysis using MS intensities from raw data and generates P-values for each metabolite. Obtained features were identified based on exact mass and relative retention time. They were then searched against a database of accurate masses taken from KEGG [323], HMDB [191, 242] and Metlin [224] using an Excel-based macro written 'in house' at Strathclyde University. Retention times of some metabolites were matched to retention times of 82 standards, which were run in the same experiment. All masses were within 1.5 ppm of the exact molecular formula.

## 2.17 CD-spectroscopy

All CD (Circular Dichroism)-spectroscopy experiments were performed at the Scottish Circular Dichroism Facility (University of Glasgow). The system contained stopped flow instrument (Applied Photophysics SX.17MV, SK.1E and CD.1C components) allowing recording in absorbance, fluorescence and CD modes, a JASCO J-810 spectropolarimeter with peltier device and automatic titrator system and biologic SFM20 stopped flow instrument. The aperture used was a Jasco J-810 spectropolarimeter (Jasco UK LTD), equipped with an additional photomultiplier, which allows data collection over the wavelength range from 180 nm to 1000 nm.

CD spectra were recorded in a cell of 0.5 cm path length at 20°C. The readings were obtained over the wavelength range from 350 nm to 204 nm, data pitch of 0.2 nm, bandwidth of 1 nm, response of 1 sec, standard sensitivity and scanning speed of 20 nm/min.

## 2.18 Bioinformatics tools for *CG30016* and *urate oxidase* sequence analysis

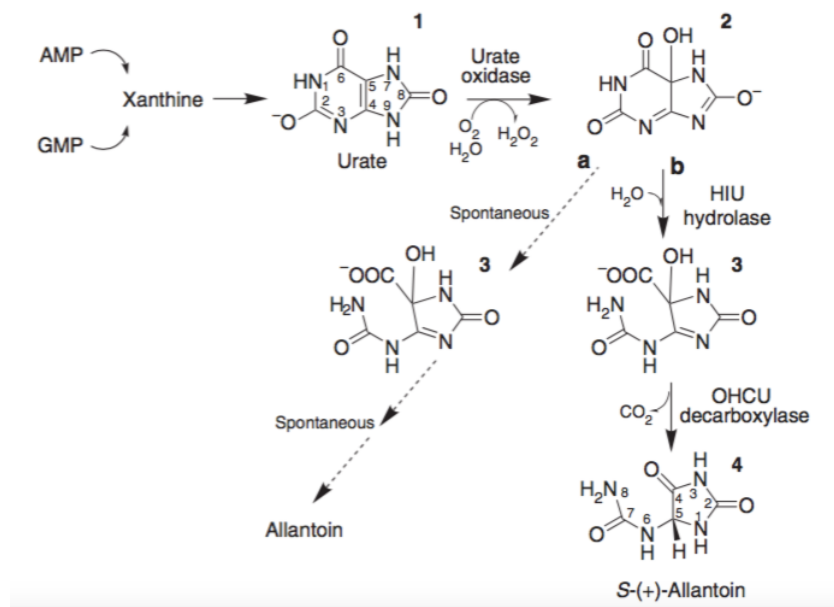
All bioinformatics analyses were performed using publicly available tools. *Drosophila melanogaster* protein sequences were obtained from FlyBase (<http://flybase.org>). Sequences of other species were from UniProt (<http://www.uniprot.org>). Protein BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) was performed with expected threshold of 100.

Sequence alignments were carried out using blastp suite-2 sequences. Pfam: protein family database (<http://pfam.xfam.org/search/sequence>) was used to determine protein family signatures. Intracellular protein localisation was predicted using PSORT II Prediction Tool (<http://psort.hgc.jp/cgi-bin/runpsort.pl>).

## 3 Urate degradation pathway

### 3.1 Introduction

Purine metabolism is one of the key metabolic pathways involved in synthesis and degradation of nucleic acids and production of important metabolites, such as ATP (Adenosine triphosphate) [354]. Excessive purines are broken down to urate, which can be secreted, stored or converted to allantoin via the urate degradation pathway. Urate degradation pathway involves three reactions that can occur spontaneously or catalysed by three distinct enzymes. The spontaneous reaction is much slower (several hours) than the enzymatic one and gives rise to racemic allantoin (R-allantoin) [355, 356]. Enzymatic degradation of urate, on the other hand, involves its stereo-specific conversion to dextrorotary allantoin, S-(+)-allantoin on a time scale of seconds. There are three enzymes responsible for this conversion: urate oxidase (UO), 5-hydroxyisourate (5-HIU) hydrolase and 2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline (OHCU) decarboxylase [102, 355, 357] (Figure 3-1).



**Fig 3-1 Formation and breakdown of urate.** Degradation of purines results in the production of xanthine, which gets oxidised by xanthine dehydrogenase to urate (1). Urate is subsequently converted by urate oxidase to 5-HIU (2). (a) 5-HIU is either spontaneously degraded to OHCU (3) and R-allantoin, which takes several, or (b) converted by 5-HIU hydrolase to OHCU and then by OHCU decarboxylase to S-allantoin (4). The enzymatic conversion of urate to S-allantoin takes several seconds. Figure adopted from [102]. Urate oxidase catalyses the conversion of urate to metastable 5-hydroxyisourate [356,

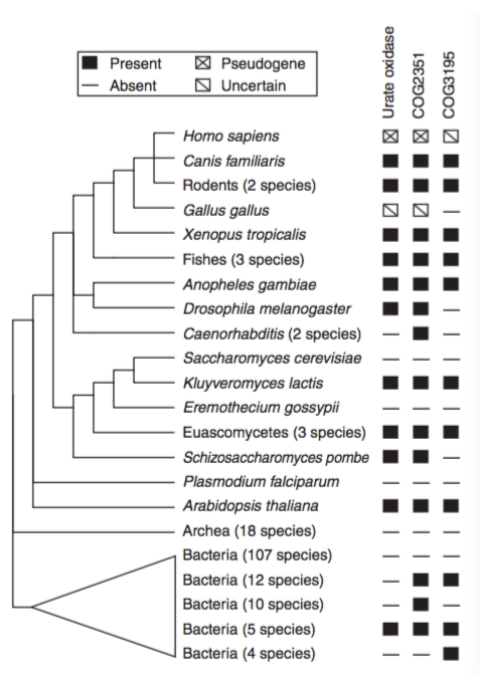
358]. 5-HIU hydrolase is responsible for stereo-specific conversion of 5-HIU to OHCU [359], which is also metastable. OHCU then gets broken down to stable S-allantoin by OHCU decarboxylase. S-allantoin can be converted to urea via two enzymatic reactions, catalysed by allantoinase and allantoicase.

### 3.1.1 Evolution of urate degradation pathway

Purine metabolism has been proposed as a candidate for one of the most ancient pathways on the planet [360]. The activity of urate degradation enzymes has been lost and gained several times during evolution (at least five independent events), with parallel loss or gain of activity of all enzymes [102, 361]. As a result, humans and other hominoids, as well as birds, some reptiles, Dalmatian dogs and several bacterial species, are not able to degrade urate and are prone to hyperuricemia [118, 119].

In humans and other hominoids urate degradation enzymes are present as pseudogenes, i.e. they are present in the genome but do not produce functional proteins [362].

Inactivation of the UO gene occurred in a primate ancestor around 15 Myr (million years) ago, which resulted in their inability to degrade urate [361]. Moreover, human *5-HIU hydrolase (5-HIUH)* gene has been inactivated by several independent mutations, whereas the expression of OHCU decarboxylase gene is absent [102]. In several bacterial species urate degradation enzymes are not present in the genome, whereas in other species only one or two of the genes are present. However, wherever UO is present, at least one of the downstream enzymes is also present (Figure 3-2). This is surprising considering that 5-HIU and OHCU can be converted spontaneously to R-allantoin. One of the reasons for the maintenance of 5-HIU hydrolase and OHCU decarboxylase could be the reaction rate. The enzymatic reaction is much faster than the spontaneous reaction. Moreover, the spontaneous reaction results in the formation of R-allantoin, which cannot be used as substrate for allantoinase [363, 364] and produce urea. Moreover, only S-allantoin has been observed in living cells [362]. R-allantoin can be converted to S-allantoin in a separate reaction catalysed by racemase. However, it takes around 10 h for this reaction to take place [363]. Once again, the enzymatic reaction would be favoured in terms of its rate. Finally, urate degradation intermediates can be further oxidised to yield reactive chemical species [365]. Hence, in order to limit their lifetime, the enzymatic reaction would be favoured.



**Fig. 3-2 Phylogenetic genome comparison of urate degradation enzymes.** Urate degradation enzymes have been lost and gained several times during evolution. They are present in most vertebrates, bacteria and plants. However, they are absent or present as pseudogenes in humans, birds and several species of bacteria. Figure adopted from [102].

### 3.1.2 *Drosophila* enzymes involved in urate degradation pathway



*Drosophila melanogaster* expresses UO and 5-HIU hydrolase and hence is able to degrade urate (Figure 3-2) [102]. There is no homologous sequence to OHCU decarboxylase in *Drosophila* genome. Whereas there are several studies confirming the structure and function of UO in the fly, there is no experimental evidence for the expression of functional 5-HIU hydrolase in the fly or its function in urate degradation pathway [366-371]. Based on sequence homology, *Drosophila* CG30016 gene codes for 5-HIU hydrolase (<http://blast.ncbi.nlm.nih.gov>). Moreover, functional relationship between UO and CG30016 has been suggested by Ramazzina *et al.* due to correlated gained or loss of activity of these two genes during evolution [102]. FlyAtlas revealed that both UO and CG30016 are enriched in MTs, which also suggests a functional relationship between the two genes [4, 6].

One possibility is that *Drosophila* gene CG30016 encodes functional 5-HIU hydrolase, which converts 5-HIU to OHCU, which spontaneously degrades to R-allantoin. Another possibility is that CG30016 catalyses both reactions, resulting in the production of S-

allantoin, or is involved in another metabolic pathway altogether.

### 3.1.2.1 *Drosophila* Urate oxidase

*Drosophila* UO has been extensively studied in terms of its structure, function and temporal regulation [366-371]. *UO* gene is located on chromosome 2 [367] and *UO* mRNA is expressed in the main segment cells of MTs [369]. As previously mentioned, UO is expressed almost exclusively in the MTs [4, 6, 368] (Figure 3-3), where it localises to the peroxisome [369]. Temporal regulation of its activity is tightly controlled. It is not detectable in L1 and L2 larval stages or pupae, whereas its expression is the highest during L3 stage and detectable in adult stages [368, 370]. Sudden decrease in UO activity right before pupation results in the accumulation of urate [367]. Research suggests that this temporal regulation is control by developmental clock in the MTs as well as haemolymph factors [371]. Moreover, there is evidence suggesting that steroid hormone 20-hydroxyecdysone represses UO activity by either limiting its transcription or increasing degradation of UO mRNA [367]. Whereas UO expression is induced by urate in some microorganisms, it is not the case in the fruit fly [370]. Moreover, its activity is inhibited by xanthine and hypoxanthine.

<div> <div>  <div> Name: <b>Urate oxidase</b> Symbol: <b>Uro</b> </div> </div> <div> Annotation Symbol: <b>CG7171</b> FlyBase ID: <b>FBgn0003961</b> Probe Set ID: <b>1625436_at</b>  </div> </div>				
Tissue	Adult Abundance	Adult Enrichment	Larval Enrichment	Larval Abundance
Head	ND	ND		
Eye	ND	ND		
Adult Brain / Larval CNS	ND	ND	[0.0]	[5]
Thoracoabdominal ganglion	ND	ND		
Crop	ND	ND		
Midgut	ND	ND	ND	ND
Hindgut	ND	ND	[0.0]	[10]
Tubule	6590	36	16	2892
Fat body	[40]	[0.2]	ND	ND
Salivary gland	[17]	[0.1]	ND	ND
Heart	[13]	[0.1]		
Trachea			ND	ND
Ovary	ND	ND		
Virgin Spermatheca	[15]	[0.1]		
Mated Spermatheca	ND	ND		
Testis	ND	ND		
Accessory glands	[12]	[0.1]		
Carcass	[7]	[0.0]	ND	ND

**Fig. 3-3 Tissues-specific expression of Urate oxidase in *Drosophila melanogaster*.** Fly UO is almost exclusively expressed in MTs, with very low expression in the fat body, salivary glands, heart, virgin spermatheca, accessory glands and carcass. Figure adopted from <http://flyatlas.gla.ac.uk>.

### 3.1.2.2 5-Hydroxyisourate hydrolase

5-Hydroxyisourate hydrolase has been studied in several species including soybean [102, 359, 372], mouse [102, 118], *Aspergillus nidulans* [373], rainbow trout [374], zebra fish [375], *Salmonella dublin* [376], *Escherichia coli* [377, 378], *Cryptococcus neoformans* [379], *Klebsiella pneumoniae* [380], South American opossum [381], the platypus [381], *Caenorhabditis elegans* [377], *Herbaspirillum seropedicae* [382], *Branchiostoma japonicum* [383] and *Bacillus subtilis* [384]. With exception of the soybean enzyme, 5-HIU hydrolase is a member of transthyretin-related protein (TRP) family. TRPs have around 30-40% sequence similarity with transthyretin, vertebrate transport protein of the thyroid hormone and retinol, but are not functionally related [377, 384, 385]. This is an example of divergent evolution, which took place around 500-700 Myr ago [386-388]. It has been proposed that during that time transthyretin arose as a result of gene duplication of 5-HIU hydrolase gene [102, 375]. As a result, transthyretin is only present in vertebrates, whereas TRPs are present in bacteria, plants and animals [389, 390]. Both proteins have a similar homotetrameric structure. Moreover, 5-HIU hydrolase active site corresponds to thyroid hormone binding site of transthyretin [376]. As opposed to transthyretin, TRPs do not bind thyroid hormone [377]. However, single or multiple mutations in TRP active site have been shown to abolish 5-HIU hydrolase activity and result in thyroid hormone binding [375]. General characteristics of TRP family are summarised in Table 3-1.

**Table 3-1 Characteristic TRP features.**

TRP feature	Description
Protein size	Around 120 amino acids
Sequence similarity to transthyretin	30-40%
C-terminal tetrapeptide sequence	Y-R/K-G-S/T - part of the active side
Conserved residues	His-14, Arg-49, His-105
Tertiary structure	Two $\beta$ -sheets and one short $\alpha$ -helix
Quaternary structure	Homo-tetramer

In terms of cellular localization, 5-HIU hydrolases can be divided into three different groups. Peroxisomal enzymes contain N-terminal peroxisomal targeting signal 2 (PTS2)

with a consensus of  $RL_{x2}L_{x2}HL$ , and are mostly found in plants and most metazoans [102, 391, 392]. Peroxisomal localisation is consistent with 5-HIU hydrolase role in urate degradation pathway, as UO also localises to the peroxisome. The second group comprises cytoplasmic enzymes, which lack any signal signature [377, 393, 394]. This is usually the case for bacterial 5-HIU hydrolases. The third group represents periplasmic proteins, which are found in enterobacteria [388].

As previously mentioned, soybean 5-HIU hydrolase is not related to TRPs and shares no sequence similarity with 5-HIU hydrolases of other species [395]. The soybean enzyme is a member of family one glycosidases, which suggests it emerged through a different evolutionary event [359, 372]. In fact, purine metabolism of leguminous plants is distinct from other species and is not well defined [119]. Moreover, there is another protein in soybean identified as TRP, which does not have 5-HIU hydrolase activity [395].

## 3.2 Results

### 3.2.1 *Drosophila melanogaster* CG30016 gene

*Drosophila* CG30016 is expressed almost exclusively in the MTs [4, 6, 368] (Figure 3-4), which is consistent with the expression of the upstream *Drosophila* enzyme, UO, and suggests functional correlation. Moreover, it exhibits features consistent with TRPs as well as 5-HIU hydrolase. It is 113 amino acids long and contains the C-terminal YRGS tetrapeptide and conserved residues characteristic to TRPs. It shares sequence homology with 5-HIU hydrolases of other species and has transthyretin and 5-HIU hydrolase domains (<http://www.ncbi.nlm.nih.gov>).



<div> <div>Symbol: <b>CG30016</b></div> <div> <div>Annotation Symbol: <b>CG30016</b></div> <div>FlyBase ID: <b>FBgn0050016</b></div> <div>Probe Set ID: <b>1626600_at</b></div> </div> </div>				
Tissue	Adult Abundance	Adult Enrichment	Larval Enrichment	Larval Abundance
Head	ND	ND		
Eye	ND	ND		
Adult Brain / Larval CNS	ND	ND	[0.1]	[11]
Thoracoabdominal ganglion	ND	ND		
Crop	ND	ND		
Midgut	ND	ND	ND	ND
Hindgut	40	0.40	[0.2]	[17]
Tubule	3945	35	28	3206
Fat body	[11]	[0.1]	ND	ND
Salivary gland	ND	ND	ND	ND
Heart	ND	ND		
Trachea			[0.1]	[8]
Ovary	ND	ND		
Virgin Spermatheca	ND	ND		
Mated Spermatheca	ND	ND		
Testis	ND	ND		
Accessory glands	ND	ND		
Carcass	[10]	[0.1]	ND	ND

**Fig. 3-4 Tissue-specific expression of *CG30016* in *Drosophila melanogaster*. *CG30016* is almost exclusively expressed in MTs, with low expression in the hindgut, fat body, carcass and larval trachea. Figure adopted from <http://flyatlas.gla.ac.uk>.**

A sequence alignment of the fly *CG30016* gene with experimentally confirmed 5-HIU hydrolases of other species was performed using protein BLAST (blastp-2 sequences). The alignment revealed 30-41% sequence similarity to 5-HIU hydrolases and presence of conserved amino acids in *CG30016* gene, specifically C-terminal YRGS, His-14, Arg-49 and His-105 (mouse numbering) (Figure 3-6). *CG30016* was also aligned with mouse and human transthyretin genes (Figure 3-5). This revealed sequence identities of 24% (E value = 4e-05) and 23% (E value = 5e-09) to human and mouse gene respectively. Finally, Pfam (<http://pfam.xfam.org/search/sequence>), the protein family database that uses multiple sequence alignment, was used to establish whether *CG30016* belongs to a known protein family. The search revealed that the *Drosophila* protein has the transthyretin family signature (HIUase/transthyretin family) and the E value was 4.8e-33. Altogether, these results show that *Drosophila CG30016* gene is related to transthyretin protein family and hence has a common ancestor with transthyretin.

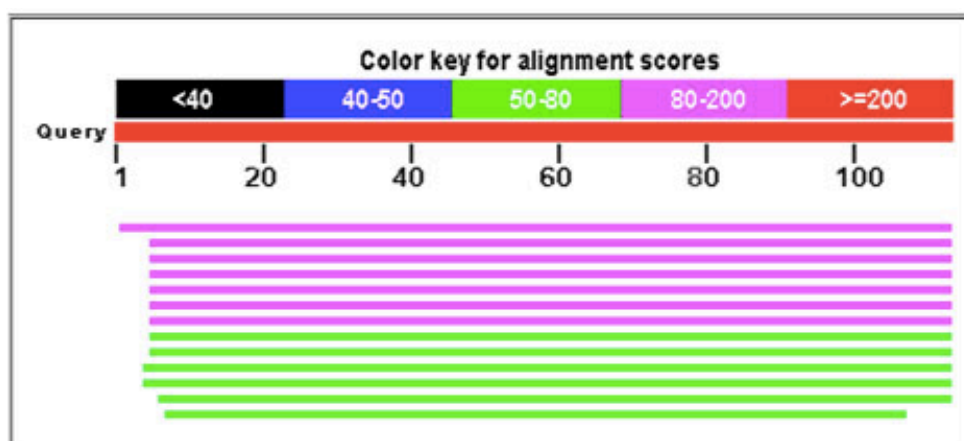
## A

Score	Expect	Method	Identities	Positives	Gaps
37.7 bits(86)	5e-09	Compositional matrix adjust.	24/103(23%)	46/103(44%)	3/103(2%)
Query 20	ILDTSVGKAAANVRVTVSRLDEIQEWRLRAAQTDADGRCL-LLEPGQFPGGIYKLTFFHV				78
	+LD G A +V V V + W + +T G L +F G+Y++				
Sbjct 41	VLDAVRGSPAVDVAVKVFKKTSEGSWEPPFASGKTAESGELHGLTTDEKFVEGVYRVELDT				100
Query 79	GAYYAERNVRTLYPAIDLIVDCSEN--QNYHIPLLLNPFPGYST				119
	+Y+ + + D++ +++ ++Y I LL+P+ YST				
Sbjct 101	KSYWKTGISPFFHEFADVFTANDSGHRHYTIAALLSPYSYST				143

## B

Score	Expect	Method	Identities	Positives	Gaps
24.3 bits(51)	4e-05	Compositional matrix adjust.	12/51(24%)	26/51(50%)	2/51(3%)
Query 1	IYKLTFFHVGAYYAERNVRTLYPAIDLIVDCSEN--QNYHIPLLLNPFPGYST				49
	IYK+ +Y+ + + +++ +++ + Y I LL+P+ YST				
Sbjct 88	IYKVEIDTKSYWKALGISPFHEHAENVFTANDSGPRRYTIAALLSPYSYST				138

Fig. 3-5 Sequence alignment of *CG30016* with mouse (A) and human (B) transthyretin genes. *CG30016* has (A) 23% sequence identities with mouse transthyretin, E value = 5e-09 and (B) 24% sequence identity with human transthyretin, E value = 4e-05.



		Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">Mouse</a>	95.1	95.1	99%	2e-30	40%	Query_94158
<input type="checkbox"/>	<a href="#">Platypus</a>	91.7	91.7	95%	4e-29	41%	Query_94166
<input type="checkbox"/>	<a href="#">Opossum</a>	90.1	90.1	95%	1e-28	41%	Query_94167
<input type="checkbox"/>	<a href="#">Opossum</a>	90.1	90.1	95%	1e-28	41%	Query_94155
<input type="checkbox"/>	<a href="#">Klebsiella pneumoniae</a>	86.3	86.3	95%	6e-27	38%	Query_94165
<input type="checkbox"/>	<a href="#">Escherichia coli</a>	84.7	84.7	95%	3e-26	40%	Query_94159
<input type="checkbox"/>	<a href="#">Zebrafish</a>	81.6	81.6	95%	5e-25	37%	Query_94160
<input type="checkbox"/>	<a href="#">Salmonella</a>	79.7	79.7	95%	3e-24	38%	Query_94156
<input type="checkbox"/>	<a href="#">Salmonella</a>	79.7	79.7	95%	3e-24	38%	Query_94162
<input type="checkbox"/>	<a href="#">Bacillus subtilis</a>	76.3	76.3	96%	4e-23	39%	Query_94157
<input type="checkbox"/>	<a href="#">Pseudomonas</a>	73.2	73.2	96%	9e-22	35%	Query_94163
<input type="checkbox"/>	<a href="#">Caenorhabditis elegans</a>	73.2	73.2	94%	1e-21	33%	Query_94161
<input type="checkbox"/>	<a href="#">Rainbow trout</a>	63.5	63.5	88%	3e-18	30%	Query_94164

**Fig. 3-6 Similarity metrics of *CG30016* gene with 5-HIU hydrolases from other species, revealing 30-41% sequence identity. E-values for each species are also provided.**

Finally, *CG30016* gene was aligned with soybean 5-HIU hydrolase in order to establish whether the fly protein was related to TRPs or family one glycosidases. The alignment, showed no sequence similarity between the two genes with E value of 5.4 (Figure 3-7), suggesting they are not related. Moreover, protein BLAST search revealed that the soybean enzyme had 33% sequence similarity with other *Drosophila* gene, which is not related to urate degradation pathway.

Score	Expect	Method	Identities	Positives	Gaps
13.1 bits(22)	5.4	Compositional matrix adjust.	4/7(57%)	6/7(85%)	0/7(0%)
Query	1	MDARKFS	7		
		+DA +FS			
Sbjct	105	LDAYRFS	111		

**Fig. 3-7** Sequence alignment of *CG30016* with soybean 5-hydroxyisourate hydrolase. Very little sequence similarity was revealed between the two proteins with the E value of 5.4, suggesting that the proteins are not related.

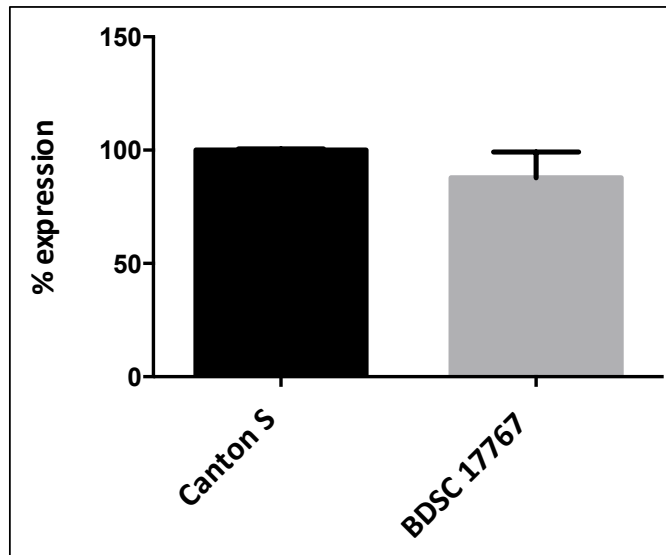
Based on its sequence, it is not clear where *CG30016* localises to in the cell. It lacks N-terminal peroxisomal targeting signal 2 (PTS2), which suggests it is not a peroxisomal enzyme. Moreover, intracellular localization was predicted based on the presence of signal peptides using PSORT II Prediction Tool (<http://psort.hgc.jp/cgi-bin/runpsort.pl>). This revealed that *CG30016* is most likely a cytoplasmic protein (Figure 3-8). However, absence of signaling peptide does not confirm or refute localization of a protein.

```
52.2 %: cytoplasmic
21.7 %: cytoskeletal
17.4 %: nuclear
 4.3 %: peroxisomal
 4.3 %: mitochondrial
```

**Fig. 3-8** Prediction of intracellular localisation of *CG30016*. Localisation of *CG30016* is most likely cytoplasmic.

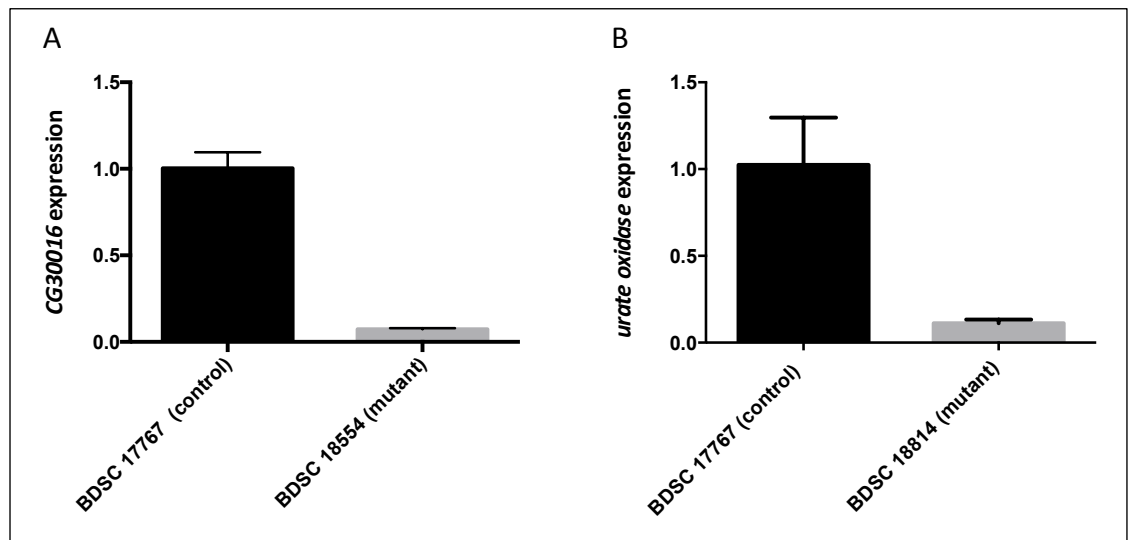
### 3.2.2 *Drosophila CG30016* and *Urate oxidase* mutants

In order to study the function of *Drosophila CG30016* and *uro* genes, mutant flies thought to lack *CG30016* or *UO* activity were obtained from Bloomington Stock Centre (see Table 2-1 for details of fly lines ordered). In order to validate deletion lines, qPCR was performed using *CG30016* and *UO* primers (see Appendix II for primer sequences). Control line (BDSC 1776) was selected with the same genetic background as the mutant lines. BDSC 17767 flies have the same P element (PiggyBac) as mutant flies, inserted in the intron region of *hbs* gene without affecting its expression. This was confirmed using qPCR and by comparing *hbs* expression in Canton S wild-type flies and BDSC 17767 flies (Figure 3-9). Absolute percentage *hbs* expression in BDSC 17767 flies was  $87.75 \pm 5.7$ , and there was no statistically significant difference in the expression between BDSC 17767 and Canton S (CS) flies.



**Fig. 3-9 Expression of *hbs* gene in BDSC 17767 line (control flies).** A qPCR was run and *hbs* expression in BDSC 17767 flies was compared to Canton S flies. There was no significant change in the expression levels of *hbs* between these two lines, validating BDSC 17767 as a good control for subsequent experiments.

Having validated BDSC 17767 line as a control with the same genetic background and not disrupting expression of any *Drosophila* genes, qPCR was performed to confirm BDSC 18554 and BDSC 18814 deletion lines using RNA extracted from adult Malpighian tubules. The expression levels were monitored and represented as values of 0.0 – 1.0, with 1.0 corresponding to 100% expression. Absolute knockdown of *CG30016* for BDSC 18554 was  $0.073 \pm 0.004$  (t-test,  $P < 0.0001$ ), whereas knockdown of *UO* for BDSC 18814 was  $0.11 \pm 0.013$  (t-test  $< 0.005$ ). The results are illustrated in Figure 3-10.



**Fig. 3-10 QPCR validation of mutant flies BDSC 18554 and BDSC 18814; gene expression was compared to BDSC 17767 control flies and expressed as 0.0 – 1.0, with 1.0 corresponding to 100% expression. (A) Expression of *CG30016* gene in BDSC 18554 line**

showing over 99% knockout of *CG30016* (t-test,  $P < 0.0001$ ) (B) Expression of *UO* gene in BDSC 18814 line confirming over 85% knockout of *urate oxidase* (t-test,  $P < 0.005$ ).

### 3.2.3 *CG30016* and *Urate oxidase* expression at different developmental stages in *Drosophila* larvae

As previously mentioned, *UO* is not expressed in *Drosophila* larvae before L3 stage. In order to confirm that this is also the case in studied fly lines, PCR was performed using DNA templates from different larval stages (L1-L3). Results are illustrated in Figure 3-11 and confirm that *UO* is not expressed in larvae before L3E stage in control flies. Moreover, it is not expressed in *CG30016* mutant until L3 stage, where its expression appears to be reduced compared to the control line. This suggests that *CG30016* knockdown in BDSC 18554 line, also affects the expression of *UO*.

To test whether *CG30016* has the same expression pattern during larval development, PCR was also performed using *CG30016* primers. As opposed to *UO* expression, *CG30016* is present at all larval stages. However, its expression appears to be the highest in L3 larvae. Moreover, there is no difference in *CG30016* expression levels between the control and *UO* mutant. This suggests that *UO* knockdown has no effect on *CG30016* expression. Tubulin expression was monitored to provide a loading control.

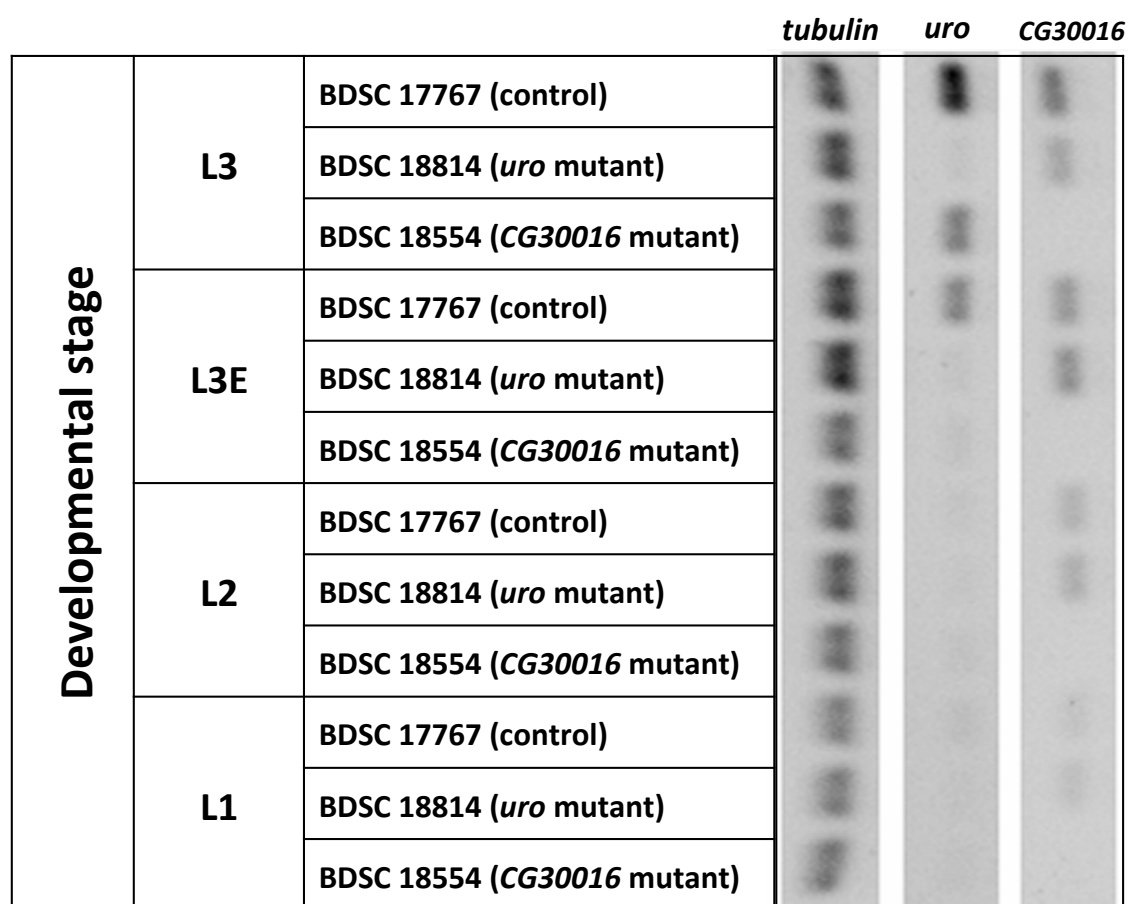


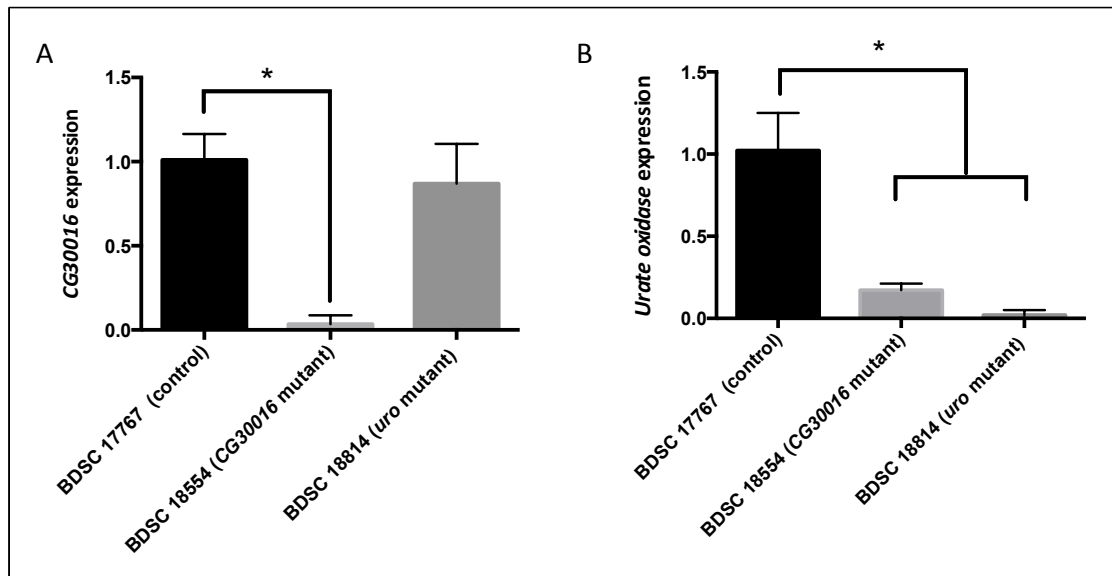
Fig. 3-11 Expression of *UO* and *CG30016* genes throughout larval development. PCR was performed and gene expression was monitored in control flies (BDSC 17767), *CG30016* deletion flies (BDSC 18554) and *UO* deletion flies (BDSC 18814). *Tubulin* expression was monitored as a loading control. L1 –L3 represent different larval developmental stages (see Table 2-2 for details).

### 3.2.4 Expression of *CG30016* and *Urate oxidase* in mutant L3 larvae

Following from the PCR results and the observation that both *CG30016* and *UO* are expressed at high levels in L3 larvae, qPCR was performed to validate these results in whole L3 larvae. Absolute knockdown value of *CG30016* for BDSC 18554 was  $0.034 \pm 0.05$  (t-test,  $P < 0.0001$ ), whereas for BDSC 18814, it was  $0.87 \pm 0.08$  (not statistically significant) (Figure 3-12 A).

Expression values of *UO* were  $0.17 \pm 0.01$  and  $0.01 \pm 0.01$  for BDSC 18554 and BDSC 18814 respectively with  $P < 0.0001$  (Figure 3-12 B).

These results confirm that both *UO* and *CG30016* are expressed in control L3 larvae. Moreover, the expression of *UO* is affected by the knockout of *CG30016*, whereas the opposite is not the case. Knocking out of *CG30016* significantly reduced the expression of



**Fig. 3-12 Expression of *UO* and *CG30016* genes in L3 mutant larvae. (A) Expression of *CG30016* in *UO* and *CG30016* deletion larvae compared to BDSC 17767 larvae, confirming over 90% knockout of *CG30016* (t-test,  $P < 0.0001$ ) in BDSC 18554 line. No significant change in expression of *CG30016* was observed in BDSC 18814 larvae (B) Expression of *UO* in *UO* and *CG30016* deletion flies, confirming over 90% knockout of *uro* (t-test,  $P < 0.0001$ ) in BDSC 18814 line, as well as over 80% knockout of *uro* (t-test,  $P < 0.0001$ ) in BDSC 18554 line.**

### 3.3 Discussion

The results suggest that *Drosophila* gene *CG30016* codes for 5-HIU hydrolase. Sequence analysis of *CG30016* revealed that it is related to TRPs and homologous to 5-HIU hydrolases of other species. However, it is not related to soybean 5-HIU hydrolase, which is thought to be a member of family one glycosidases [359]. Interestingly, another *Drosophila* gene, *CG33964*, has been previously reported as a member of TRP family, based on its sequence [377]. However, looking at our results, it is clear that *CG30016* is a more likely candidate for a TRP due to higher sequence similarity, presence of conserved residues and its enrichment in MTs.

TRPs and transthyretin are an interesting case of divergent evolution. Transthyretin domain evolved from 5-HIU hydrolase [375, 388]. Even though the two types of protein have a completely different function, and transthyretin is only present in vertebrates, they exhibit high sequence homology and have similar structure and functional domains. 5-HU hydrolase does not bind thyroid hormone [377]. However, single or multiple nucleotide substitutions in the active site have been shown to result in thyroid hormone



binding [375]. Insects and other invertebrates do not produce thyroid hormone. Hence, it is interesting to see that they produce a protein so closely related to transthyretin, which has no functional significance in these species.

Based on its sequence, it is not clear which cellular compartment *CG30016* localises to. It does not have any known signal peptides and hence has been hypothesised to localise to the cytoplasm. Interestingly, in rainbow trout, 5-HIU hydrolase localizes to the cytoplasm, even though it contains PTS2 signal [374]. This suggests that 5-HIU hydrolase localization might depend on various different factors rather than solely the presence of a signal sequence. Moreover, several peroxisomal proteins have been identified, which did not have a known signal peptide, suggesting a unique means of targeting[394]. Hence, it requires further studies to determine the intracellular localization of the protein encoded by *CG30016*.

Further experimental evidence is required to confirm the role of *CG30016* as 5-hydroxyisourate hydrolase. Here, we proposed the reverse genetics approach, where the function of the gene is investigated by studying the effects of its knockout. *CG30016* and *UO* knockout flies were obtained from stock centres and validated. Two controls were also selected: wild-type Canton S flies as well as flies with the same genetic background as mutants.

Finally, the results confirmed that *UO* is only expressed in L3 and adult stages of *Drosophila*. Hence, all our studies were conducted using either L3 larvae or adult flies. We also showed that *UO* expression is affected by the knockout of *CG30016*. Expression of *UO* was significantly reduced in *CG30016* knockout larvae. This could be a result of a feedback mechanism, where *CG30016* or products of a reaction it catalyses are required for the transcription of *UO*, or the substrate of a reaction it catalyses inhibits *UO* transcription. For example, whenever *CG30016* is absent, it results in built up of 5-HIU. In turn, there is no need for the conversion of urate to 5-HIU, and *UO* transcription is repressed. To test this, further experimental evidence is required. The correlation between expression of *UO* and *CG30016* also confirms that the two genes are functionally related and strengthens our hypothesis that *CG30016* is involved in urate degradation pathway as 5-HIU hydrolase.

## **4 Phenotypic manifestations in *Drosophila melanogaster* mutants of urate degradation pathway**

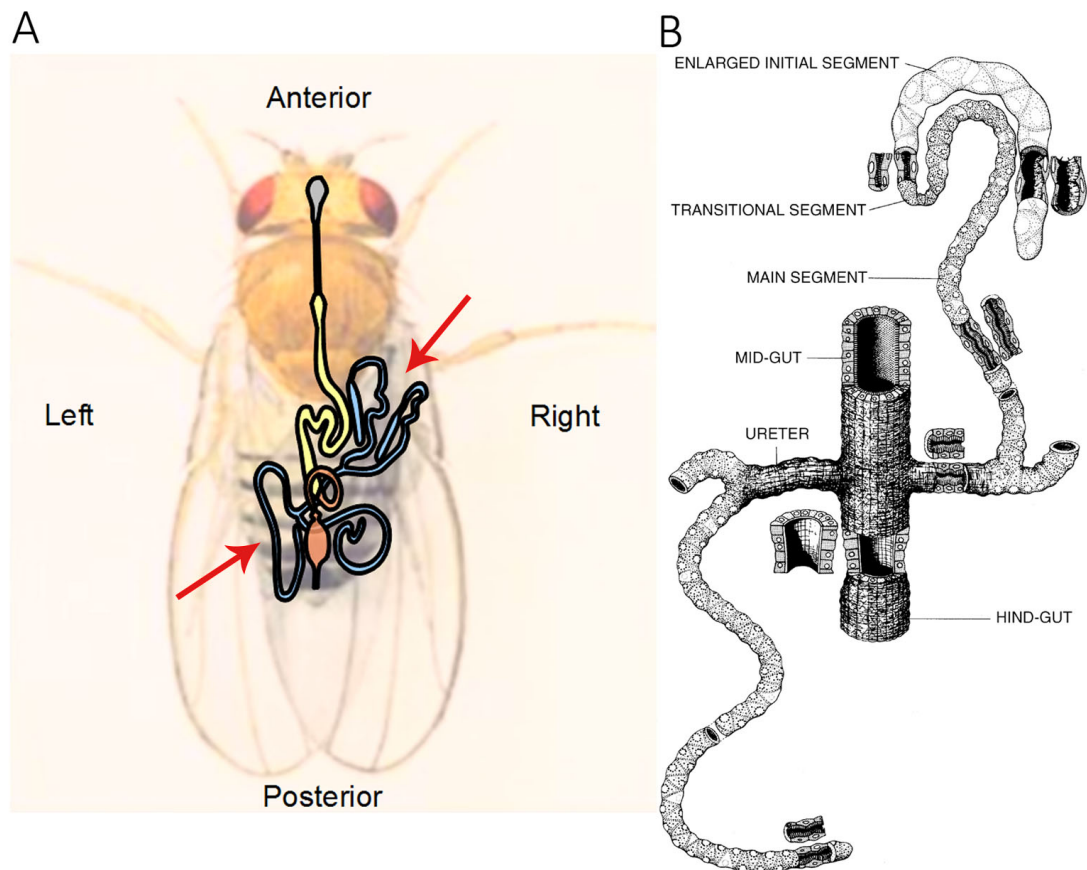
### **4.1 Introduction**

Manipulations of *Drosophila* genes have been shown to result in various phenotypic manifestations including MTs phenotypes [65-68], bloated abdomen [69], antennae and bristle phenotype [70], altered fluid secretion rate [71-74], and altered longevity and survival in different stress conditions [75, 76]. In this chapter, the effects of *CG30016* and *UO* knockout mutants on *Drosophila melanogaster* phenotype are evaluated.

#### **4.1.1 *Drosophila melanogaster* Malpighian tubules**

Fruit flies have two pairs of Malpighian tubules (anterior and posterior tubules), each pair connected to the hindgut of the alimentary canal via a ureter (Figure 4-1). Two tubule pairs are asymmetrical and are morphologically and functionally distinct [352, 396]. Anterior tubules have an enlarged initial segment, which is responsible for  $\text{Ca}^{2+}$  titreing. Posterior tubules, on the other hand, have a smaller initial segment. In terms of their function, anterior tubules are involved in immune defence and handling of calcium ions. Posterior tubules play a role in ammonia generation [396]. *Drosophila* tubules are composed of around 150 specialised excretory cells [397]. Moreover, the main segment of the tubules consists of two cell-types: principal and stellate cells [352, 398]. They differ in terms of morphology and function; principal cells are responsible for electrogenic cation transport, whereas stellate cells play a role in anion conductance [69].

Main physiological function of MTs is generation and reabsorption of primary urine, regulation of ions and acid-bases, as well as excretion of waste and toxins. Moreover, tubules have been shown to play a role in immunity, osmoregulation, metabolism and xenobiotic excretion [397, 399].



**Fig 4-1 *Drosophila melanogaster* morphology. (A) Two pairs of Malpighian tubules (blue, marked with arrows) are connected to the hindgut (orange) via a ureter. Tubule pairs are asymmetrical; the anterior tubule (right) is wrapped around the midgut; the left pair is directed posteriorly. Adopted from [396] (B) Morphology of the adult *Drosophila melanogaster* Malpighian tubules. Anterior tubule (top) has an enlarged initial segment and the posterior tubule (bottom) has a smaller initial segment. Adopted from [36].**

#### 4.1.1.1 Waste and fluid secretion

Terrestrial insects, including *Drosophila melanogaster*, eliminate nitrogenous waste products as urate [68]. This strategy has been developed to preserve water in these species. Soluble urate is transported to the MT lumen, where it gets acidified and forms uric acid crystals [400]. The crystals can be visualised due to their birefringent property under polarized light [401]. UO can further degrade urate as described in Section 3.1.2.1. However, due to temporal regulation of its activity (UO is not detectable in L1 and L2 larval stages or pupae [368, 370]), urate crystals can be observed in the tubule lumen of wild type pupae, L1, L2 and L3E larvae but not L3 larvae and adult flies.

*Drosophila* tubules have been used as a model of insect fluid secretion [352]. They are a remarkable transporting epithelium; with the highest per-cell transport rate known in nature [69]. Moreover, it is possible to study fluid secretion in dissected fly tubules for over 5 h from dissection, which makes the tissue particularly suitable as a model for fluid

secretion [352]. Previous studies showed that different factors, including gene manipulation, can influence fluid secretion rate or even arrest the process altogether [71-74].

#### **4.1.1.2 Tubule phenotypes**

Manipulation of genes enriched in *Drosophila* Malpighian tubules has been shown to result in various MTs phenotypes. Bloated tubule phenotype has been reported in fly mutants for *blot* gene [65]. This was a result of morphological distortion of the apical membrane of MT cells in these flies and the tubules had bloated appearance. Similar bloated phenotype was observed in *Drosophila maroon-like* mutants [103]. Moreover, *rosy* knockout in flies resulted in xanthine oxidase/dehydrogenase inactivation (an enzyme in purine metabolism directly upstream from UO) and accumulation of xanthine, 2- amino-4-hydroxypteridine and hypoxanthine [62, 402]. Hypoxanthine enrichment led to calculi formation in the MTs resulting in their bloating and distortion [104, 403].

Another phenotype observed in *Drosophila* MTs, is urate crystals accumulation in L3 larvae and adult flies. Moreover, accumulation of urate crystals has been reported in *UO* mutant flies [404]. On the other hand, transparent tubule phenotype has also been reported in *Drosophila melanogaster* [67]. It has been associated with mutations in V-ATPase (*vha55*) and related proteins [67, 68, 405]. These mutants fail to accumulate urate crystals in embryos, when their presence is expected. The phenotype has been described as the loss of birefringence in the tubule lumen [406] and has been associated with urinary acidification defect [67, 68, 405, 407]. Moreover, further research has shown that a mutation in the human B<sub>1</sub> subunit of the V-ATPase results in renal tubular acidosis [408]. This observation of urate crystals presence/lack formation is possible due to tissue's transparency.

Several other tubule phenotypes were previously described [66]. These were mostly related to tubule development and altered morphology caused by disruption of transcription regulation. For example, *rib* and *raw* mutants are thought to alter MT shape due to interruption of early stages of tubule development [409]. Moreover, *wingless* mutant exhibits wrong number of MTs [410].

#### **4.1.1.3 Tubules in biomedicine**

As described in Chapter 1.2 *Drosophila melanogaster* has been extensively applied as a model for studying human physiology and disease. Even though *Drosophila* tubules differ

structurally from mammalian kidneys, they are functionally analogous [63, 397]. Both human kidney and fly tubules excrete waste products, produce urine and have similar transporters. Moreover, FlyAtlas revealed that a large number of genes enriched in MTs are homologs of human disease genes involved in kidney disease and IEMs (described in more detail in Chapter 1.3.1) [6, 397, 411], for example vacuolar-type-ATPases (V-ATPases), Na<sup>+</sup>/K<sup>+</sup>-ATPase, aquaporins and several ion channels and transporters [63]. Some of these genes have been extensively studied and *Drosophila* disease models experimentally confirmed [62, 63, 99-101, 397, 412-416] (Table 4-1). For example, kidney stone formation has been recently studied in fly tubules. This is possible due to MTs being transparent, which allows real-time observation of stone formation and growth of oxalate crystals [417, 418]. Finally, MTs can be easily dissected and remain physiologically active for several hours after dissection. This makes the tissue particularly suitable for biomedical research.

**Table 4-1 Human kidney diseases modelled in *Drosophila* and fly genes homologous to human genes associated with the disease.**

<b>Human kidney disease</b>	<b><i>Drosophila</i> gene</b>
Xanthinuria type I	<i>rosy</i>
Xanthinuria type II	<i>maroon-like</i>
Bartter syndrome types 3 and 4	<i>CG31116</i>
Bartter syndrome type 2	<i>ir</i>
Colchicine resistance	<i>CG10226</i>
Renal tubular acidosis	<i>vha55</i>

#### **4.1.2 Physiological role of urate**

As described in Section 3.1.1, the ability to degrade urate has been lost and gained several times during evolution [102, 361]. As a result, urate is the final product of purine metabolism in humans and other hominoids, as well as birds, some reptiles, Dalmatian dogs and several bacterial species [118, 119]. Fruit flies, on the other hand, as well as other species, can degrade urate to allantoin and then to urea [102]. Due to inability to degrade urate, humans are susceptible to hyperuricemia [118, 119]. However, it has been

proposed that increased concentration of serum urate has an evolutionary advantage [419-421]. On the other hand, urate has also been reported to induce oxidative stress in certain conditions and lead to disease [111, 113-115, 422-427].

#### **4.1.2.1 Benefits of uric acid**

According to the oxidative stress theory of aging, free radicals lead to cellular damage, which if unrepaired, results in aging and death [428, 429]. Moreover, ROS have been suggested to contribute to cancer and cardiovascular disease. Uric acid has been shown to act as a powerful antioxidant [419, 421] and a scavenger of reactive oxygen species [119, 420, 430, 431]. Its anti-oxidative properties prevent oxidative damage of macromolecules [421, 432]. As a result, uric acid has been associated with greater longevity of invertebrates [433] and mammals [434], including primates [420]. Previous research has also suggested that it has a preventive role in cancer and cardiovascular disease [420, 421, 435]. As a result, animals with high urate concentration in the blood stream (including humans) are more likely to have increased life-span [420].

Studies in *Drosophila melanogaster* showed that *rosy* mutants, which do not produce urate, are hypersensitive to oxidative stress (induced by paraquat), confirming urate role as an antioxidant [76, 430, 436-438]. Moreover, flies unable to synthesise urate showed a shortened life-span as compared to the wild type flies [430]. Interestingly, ROS levels have been shown to increase during aging [439]. Altogether, these studies suggest that urate acts as an antioxidant in fruit flies and as a result extends life span.

#### **4.1.2.2 Urate as an inducer of oxidative stress**

On the other hand, uric acid has been associated with increased oxidative stress and disease [111, 113-115, 422-427]. Abnormally high serum urate in humans results in hyperuricemia, which is associated with gout and insulin resistance [111-115]. Moreover, uric acid is thought to play a role in cardiovascular disease, hypertension and metabolic syndrome [422, 423].

Previous research showed that uric acid increases the production of ROS and induces oxidative damage [425, 427]. One of the explanations for this is the fact that production of urate is catalyzed by xanthine dehydrogenase. During this reaction ROS are generated as by product [117, 440, 441]. Another study suggests that urate becomes pro-oxidative when it reacts with other oxidants [423]. There are currently no studies on negative effects of urate in *Drosophila melanogaster*.

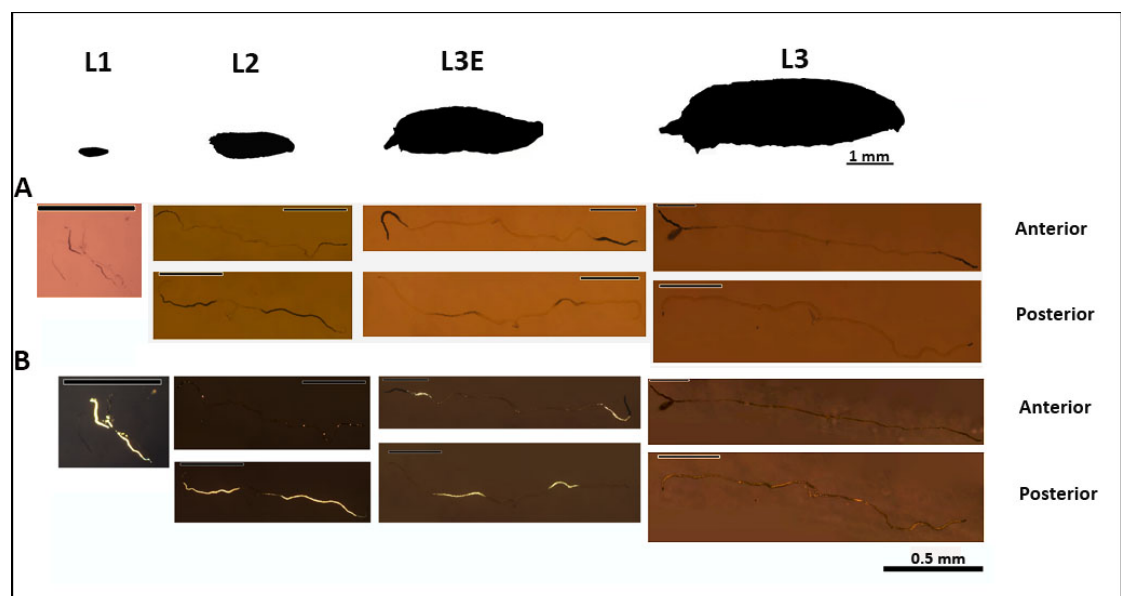
## 4.2 Results

By knocking out *CG30016* and *UO* genes in *Drosophila*, we expected to disrupt its ability to degrade urate. We expected to observe certain phenotypic manifestations of these mutations related to urate accumulation, including urate crystals formation, MT phenotype, altered survival and longevity. Since *CG30016* mutant flies are deficient in both *CG30016* and *UO*, *UO* mutant was used as a control to distinguish between phenotypes associated with *UO* knockout and not *CG30016* knockout alone.

### 4.2.1 Urate crystals

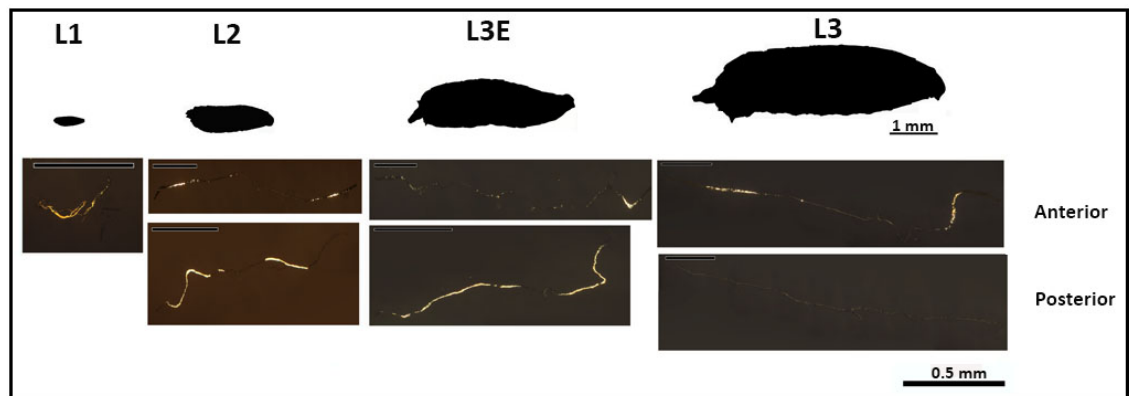
Mutant flies including BDSC 18814 and BDSC 18554 were compared to Canton S wild-type flies as well as BDSC 17767 control flies. L1-L3 larvae were dissected and their Malpighian tubules observed using light and polarised light microscopy. Polarised light microscopy allows visualisation of urate crystals due to their birefringent property [401].

Urate crystals were present in both anterior and posterior tubules of Canton S L1, L2 and L3E larvae but not in L3 larvae (Figure 4-2). This is consistent with *UO* expression at different larval stages. *UO* is not expressed during L1 and L2 stages, and it is only expressed at low levels during L3E stage. However, it is expressed in L3 larvae. Hence, no urate crystals in both posterior and anterior tubules are present at this developmental stage. Interestingly, more crystals were observed in the posterior tubule of Canton S larvae.



**Fig. 4-2 Malpighian tubules of *Drosophila* Canton S L1-L3 larvae and urate crystals. (A) Malpighian tubules of L1-L3 Canton S larvae visualised using light microscopy (B) L1-L3 tubules of Canton S flies visualised using polarised light microscopy. Urate crystals are present in L1, L2 and L3E larval stages due to the absence of *UO* activity. There are no crystals observed in L3 larvae as *UO* is expressed at this stage.**

In BDSC 17767 control larvae, urate crystals were present in both posterior and anterior tubules of L1-L3E larvae (Figure 4-3). Moreover, in contrast to Canton S flies, crystals were also observed in the anterior tubule of L3 larvae. This is not consistent with the expression patterns of *UO* at different larval stages and is somewhat puzzling. This could be due to accumulation of urate crystals in early larval stages and lower activity of *UO* in the anterior tubule of the control. Similarly to Canton S tubules, more crystals were visible in the posterior tubule of L1-L3E larvae.

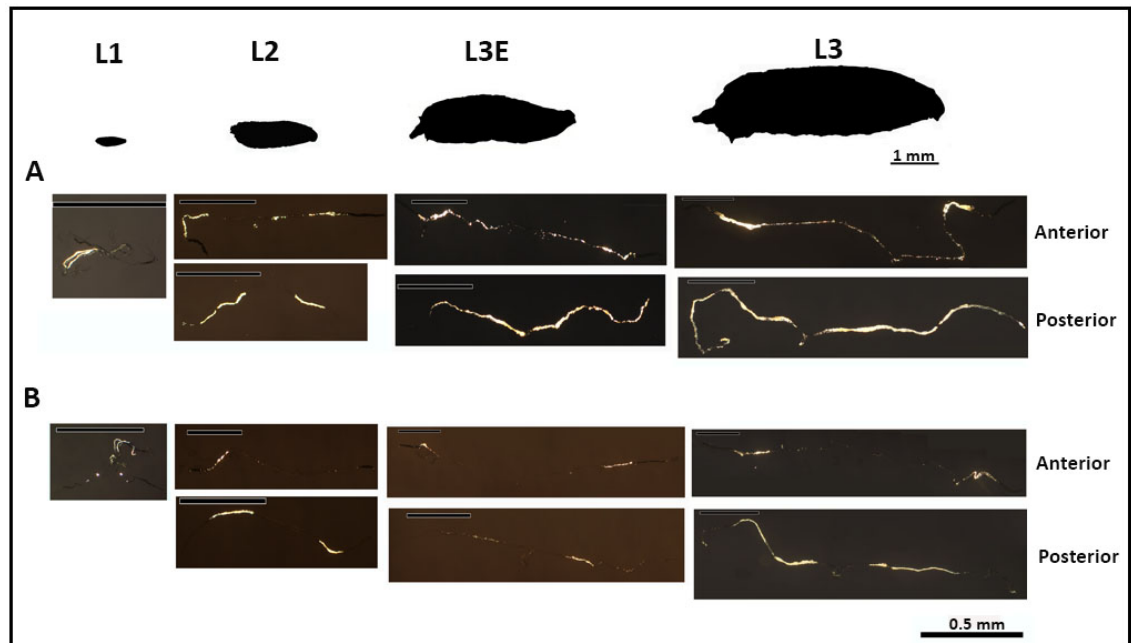


**Fig. 4-3 Urate crystals in *Drosophila* BDSC 17767 L1-L3 larval Malpighian tubules. Urate crystals are present in L1, L2 and L3E larval stages due to the absence of *UO* activity. Crystals are also observed in the anterior but not the posterior tubule of L3 larva.**

Both posterior and anterior tubules of BDSC 18814 and BDSC 18554 flies had urate crystals in their lumen at all larval stages (Figure 4-4). Moreover, markedly more crystals were observed in BDSC 18814 than in BDSC 18554 larvae. This makes sense because *UO* knockout in BDSC 18814 larvae is greater than in BDSC 18554 larvae. This suggests that low expression of *UO* in BDSC 18554 flies has an effect on urate levels in these flies. Moreover, this suggests that the knockout of *CG30016* alone does not seem to increase the accumulation of urate in mutant flies. Instead, urate crystals accumulation is the result of the *UO* knockdown in these flies. Once again, more crystals were observed in the posterior tubule than in the anterior tubule. This could be due to differential expression of *UO* in posterior and anterior tubules.



Finally, this experiment validated temporal expression of *UO* in the fruit fly. Further experiments were carried out in either L3 larvae or adult flies, as the expression of *UO* is the highest at these stages.



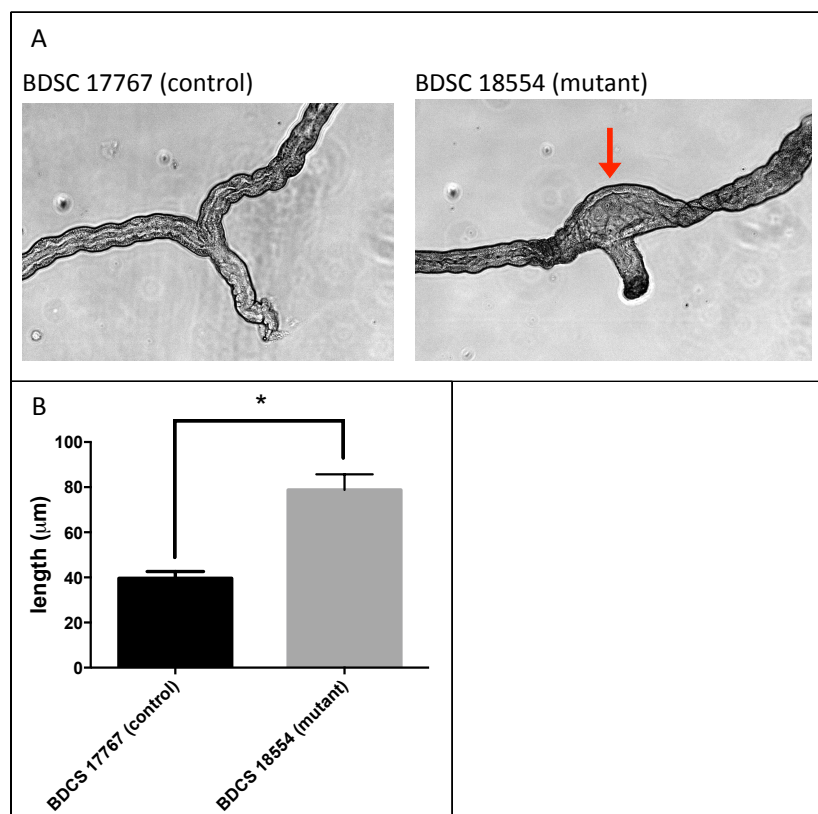
**Fig 4-4 Urate crystals in *Drosophila* Malpighian tubules.** Urate crystals are present throughout L1-L3 larval stages due to the absence of *UO* activity in these fly lines. Crystals do not disappear in L3E or L3 stages. (A) Tubules of BDSC 18814 mutant flies (*UO* knockout), lack of *UO* expression and presence of urate crystals in all larval stages. (B) Tubules of BDSC 18554 mutant flies (*CG30016* knockout with reduced *UO* expression), lack of/reduced *UO* activity and presence of urate crystals in all larval stages. Less crystals are observed in BDSC 18554 flies in comparison to BDSC 18814 flies.

#### 4.2.2 Inflated ureter phenotype

Tubule phenotype was predicted in BDSC 18554 line due to high levels of expression of *UO* and *CG30016* in MTs. Tubules of BDSC 18554 and BDSC 18814 mutant flies were visualised using light microscopy and compared to BDSC 17767 control flies. As predicted, about half of the tested BDSC 18554 flies exhibited an inflated ureter phenotype (Figure 4-5). T-test revealed that there was a statistical difference between knockout and control ureter lengths ( $p = 0.0001$ ,  $n = 100$ ). As expression of *CG30016* is much higher in male than female flies, we compared the length of ureter separately in mutant males and females. For both male and female flies the ureter length was statistically bigger in knockout flies ( $P$ -value = 0.0003). However, there was no significant difference between the size of male and female ureters in mutant flies ( $P$ -value = 0.465). It is thought that there are functional differences between anterior and posterior Malpighian tubule in

*Drosophila* because of different expression patterns so we also compared the anterior and posterior tubules of mutant flies. In both anterior and posterior tubules there was statistically significant difference in the ureter length when compared to controls (P-value = 0.0003). There were no significant differences between the size of inflation in anterior and posterior tubules in mutant flies (P-value = 0.2344). These results suggest that disruption of *CG30016* gene in *Drosophila* leads to inflated ureter irrespective of the tubule orientation or the sex of the fly.

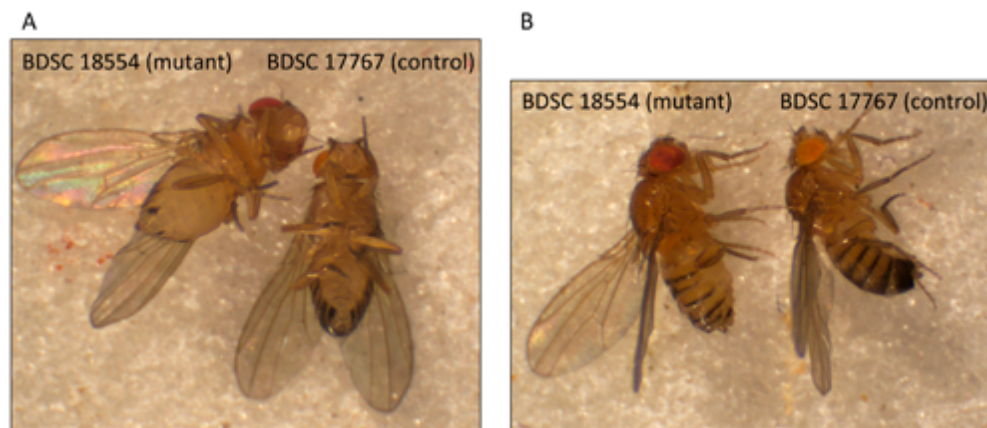
As opposed to BDSC 18554 flies, no inflated ureter was observed in BDSC 18814 flies. This suggests that the inflation is not a result of the knockout of *UO* and the accumulation of urate in mutant flies. Instead, it seems to be an effect of the knockout of *CG30016* gene alone. This could be due to accumulation of another metabolite, fluid retention in mutant ureter, inflammation or structural change of mutant tubules.



**Fig. 4-5 Inflated ureter phenotype in BDSC 18554 mutant flies. Inflated ureter phenotype was observed in the mutant flies (red arrow) but not the control flies. The size of the inflation was measured (ureter length) and compared to the size of control ureter. There was a significance difference between ureter lengths in mutant flies compared to the controls (p=0.0001).**

### 4.2.3 Inflated abdomen phenotype

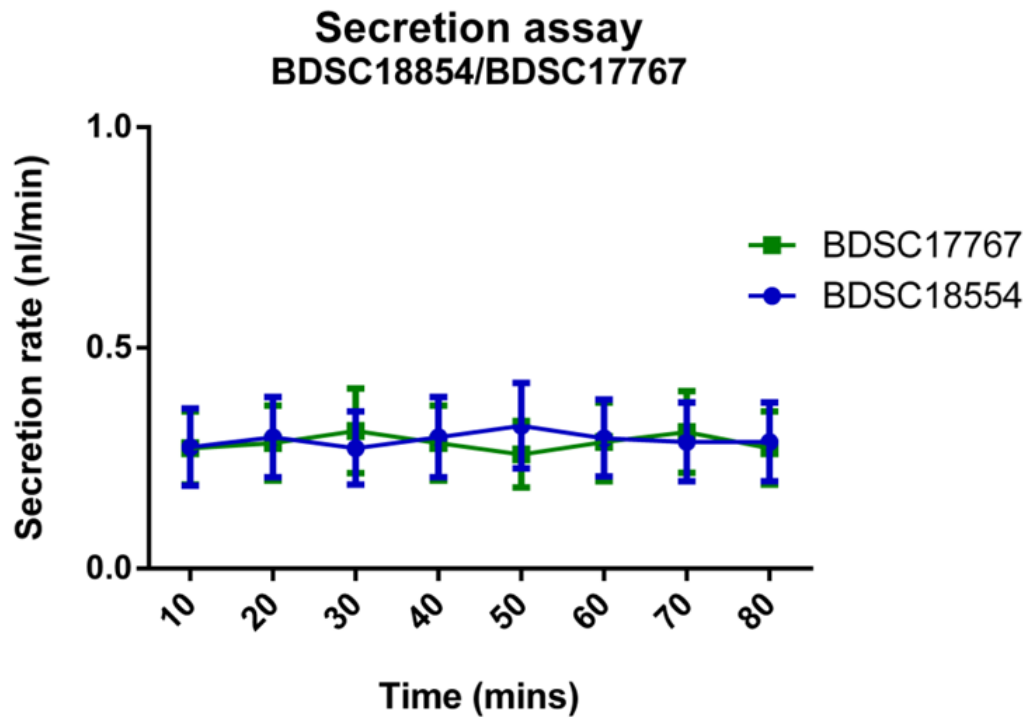
Several tissues, including gut and salivary glands, as well as whole BDSC 18554 and BDSC 18814 flies, were examined using light microscopy, in order to determine whether other phenotypes were present in these flies. There were no differences in the appearance of gut and salivary gland tissues of mutants when compared to the control. However, *CG30016* knockout flies appeared to have enlarged/bloated abdomen (Figure 4-6A) as well as distorted abdominal patterns (Figure 4-6B) in comparison to BDSC 17767 control flies. This was not the case for BDSC 18814 mutants. Once again, this suggests that this is a result of *CG30016* knockout rather than *UO* knockout. It could be a result of fluid retention, accumulation of metabolite(s), defective osmoregulation or inflammation. However, it requires further examination and statistical analysis.



**Fig. 4-6 Inflated abdomen phenotype (A) and distorted abdominal pattern (B) in BDSC 18554 mutant flies. *CG30016* knockout appear to have bloated abdomen as well as distortion of abdominal patterns.**

### 4.2.4 Tubule secretion rate

In order to establish whether *CG30016* knockout affected tubule physiology and the inflated ureter phenotype in BDSC 18554 flies arose due to fluid retention and/or reduced secretion rate, secretion assay was performed using MTs of mutant flies. Secretion rates were compared to the ones of BDSC 17767 control flies. No significant differences were observed between the rates of these two fly lines (Figure 4-7), showing that *CG30016* knockout had no effect on basal fluid secretion levels. This suggests that the inflated ureter was not a result of altered secretion rate or fluid retention related to the secretion rate of the tubule.



**Fig. 4-7 Basal secretion rate (nl/min) comparison between BDSC 18554 and BDAC 17767 flies. Tubule secretion rates were comparable in the mutant flies and the control. No statistically significant differences were observed between the two fly lines.**

#### **4.2.5 Oxidative stress survival**

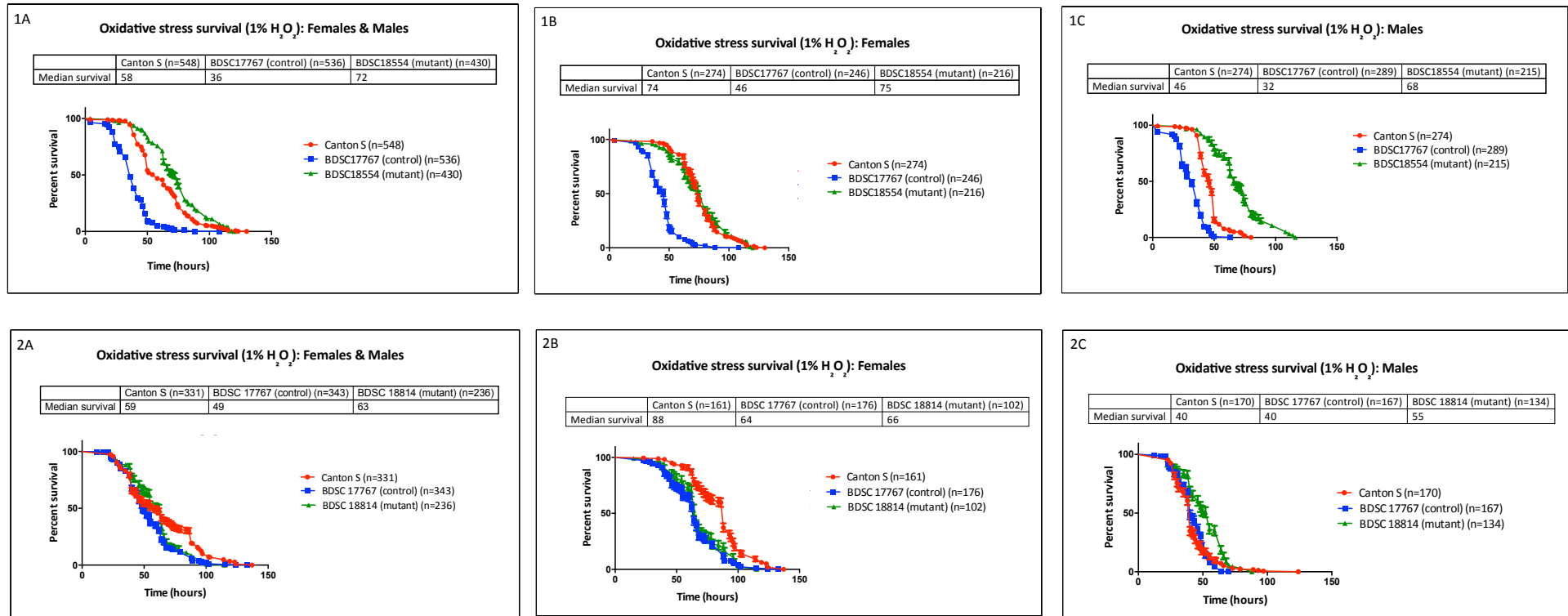
As previously mentioned, urate is a known antioxidant. Hence, urate accumulation in mutant flies was expected to prolong fly survival in oxidative stress conditions. In order to establish whether *UO* and *CG30016* knockouts have an effect on fly survival in oxidative stress conditions, survival assays were performed. Two types of oxidative stress were induced using 10 mM paraquat or 1% H<sub>2</sub>O<sub>2</sub>. Both mutant lines (BDSC 18814 and BDSC 18554) as well as BDSC 17767 control and Canton S flies were tested.

##### **4.2.5.1 Oxidative stress induced by 1% H<sub>2</sub>O<sub>2</sub>**

All female flies cope much better with oxidative stress than male flies. Moreover, BDSC 18554 flies survived significantly longer in oxidative stress conditions than BDSC 17767 control and Canton S wild-type flies and had a higher median survival than the two controls (Figure 4-8 1A). Male mutant flies cope better with oxidative stress than both BDSC 17767 controls and Canton S flies (Figure 4-8 1C). On the other hand, in female flies, *CG30016* knockout resulted in better survival and higher median survival value compared to BDSC 17767 flies but not Canton S flies (Figure 4-8 1B).

BDSC 18814 survival in response to oxidative stress is similar to both controls (Figure 4-8 2A). However, its median survival is slightly higher compared to BDSC 17767 and Canton S flies. Female Canton S flies live longer than *UO* mutants as well as BDSC 17767 controls and their median survival is higher (Figure 4-8 2B). *UO* knockout in male flies extends survival in oxidative stress conditions and their median survival is higher than the controls (Figure 4-8 2C).

In both mutants, male flies cope better than female flies with oxidative stress. The reason for this is not likely due to *CG30016* or *UO* knockout affecting males and females differently. Instead, it seems to be a result of female Canton S flies coping particularly well with oxidative stress compared to other flies (Figure 4-8 1B and 2B). Moreover, extension of survival is greater in *CG30016* mutants than in *UO* mutant flies. Since BDSC 18554 flies have decreased *UO* expression and no expression of *CG30016*, it suggests that both genes affect oxidative stress survival, and in BDSC 18554 flies their effect is synergistic.



**Fig. 4-8 Oxidative stress survival assay induced by 1% H<sub>2</sub>O<sub>2</sub>. (1A-1C) Comparison of survival in oxidative stress conditions of BDSC 18554 mutant, BDSC 17767 control and Canton S flies; (1A) Male and female flies, BDSC 18554 flies survived significantly longer than the two controls and their median survival is higher. (1B) Female flies, survival of BDSC 18554 and Canton S flies is comparable and higher than this of BDSC 17767 flies. (1C) Male flies, knockout of *CG30016* in BDSC 18554 flies results in increased survival and higher median survival in comparison to controls. (2A-2C) Comparison of survival in oxidative stress conditions of BDSC 18814 mutant, BDSC 17767 control and Canton S flies; (2A) Male and female flies, all flies respond to oxidative stress in a similar manner, median survival of *UO* mutant is slightly higher than that of the controls. (2B) Female flies, mutant and BDSC 17767 flies have comparable median survival values and both survive shorter than Canton S flies. (2C) BDSC 18814 flies cope better with oxidative stress than the controls and have higher median survival.**

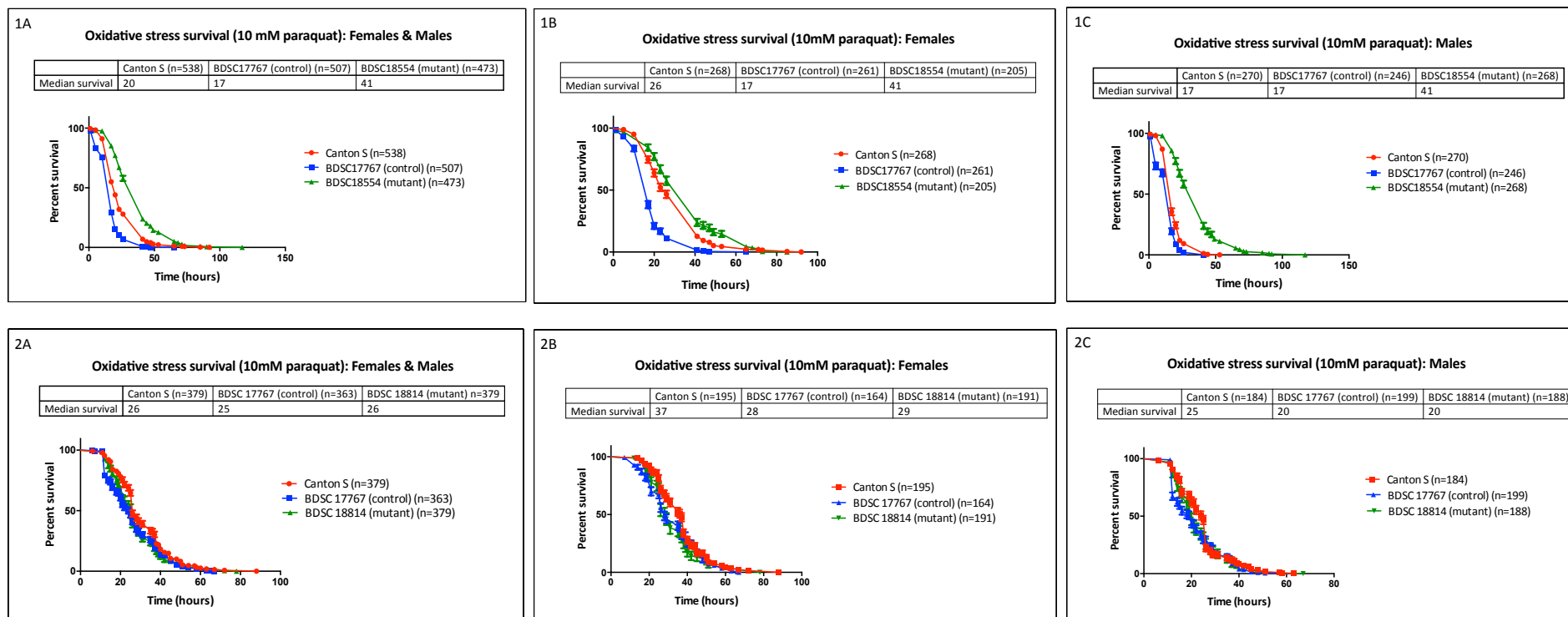
#### 4.2.5.2 Oxidative stress induced by 10 mM paraquat

Paraquat (1,1'-dimethyl-4,4'-bipyridinium dichloride) is a redox cycling compound and induces oxidative stress using a different mechanism than  $\text{H}_2\text{O}_2$ . In its reduced state it forms a stable radical species ( $\text{Pq}^+$ ). It then reacts with  $\text{O}_2$  and generates the superoxide anion [442]. In order to establish whether different types of oxidative stress have a different effect on *UO* and *CG30016* mutants, paraquat-induced oxidative stress survival assay was performed.

As was the case for  $\text{H}_2\text{O}_2$ -induced oxidative stress, BDSC 18554 flies survive significantly longer and have higher median survival than the controls (Figure 4-9 1A). As opposed to  $\text{H}_2\text{O}_2$ -induced assay, both male and female flies cope better with paraquat-induced oxidative stress than BDSC 17767 and Canton S flies and have higher median survival rates (Figure 4-9 1B and 1C). This suggests that the knockout of *CG30016* has a protective property in oxidative stress conditions; both induced by paraquat and  $\text{H}_2\text{O}_2$ . However, the effect is greater in paraquat-induced oxidative stress.

As opposed to  $\text{H}_2\text{O}_2$ -induced oxidative stress, *UO* knockout does not affect fly survival in paraquat-induced oxidative stress (Figure 4-9 2A). In both male and female flies, survival of BDSC 18814 mutant flies is comparable to BDSC 17767 control and Canton S flies (Figure 4-9 1B and 1C). However, median survival of Canton S flies (both males and females) is slightly higher than that of BDSC 17767 control and the mutant.

In agreement with  $\text{H}_2\text{O}_2$ -induced survival, the effect of *CG30016* knockout on survival is greater in males. Once again, it seems to be due to better survival of female Canton S flies than male Canton S flies.



**Fig. 4-9 Survival assay in oxidative stress conditions induced by 10 mM paraquat. (1A-1C) Survival assay using BDSC 18554 mutants, BDSC 17767 control and Cantos S wild type flies. (1A) Male and female flies, BDSC 18554 flies survived significantly longer than the two controls and their median survival is higher. (1B) Female flies, survival of BDSC 18554 is significantly higher than Canton S and BDSC 17767 flies. (1C) Male flies, knockout of *CG30016* in BDSC 18554 flies results in increased survival and higher median survival in comparison to controls. (2A-2C) Comparison of survival in oxidative stress conditions of BDSC 18814 mutant, BDSC 17767 control and Canton S flies; (2A) Male and female flies, all flies respond to oxidative stress in a similar manner and their median survival is comparable. (2B) Female flies, all flies have similar survival patterns, median survival of Canton S flies is higher than BDSC 17767 control and the mutant. (2C) Survival of all fly lines is comparable, median survival value of Canton S flies is slightly higher than the other two lines.**

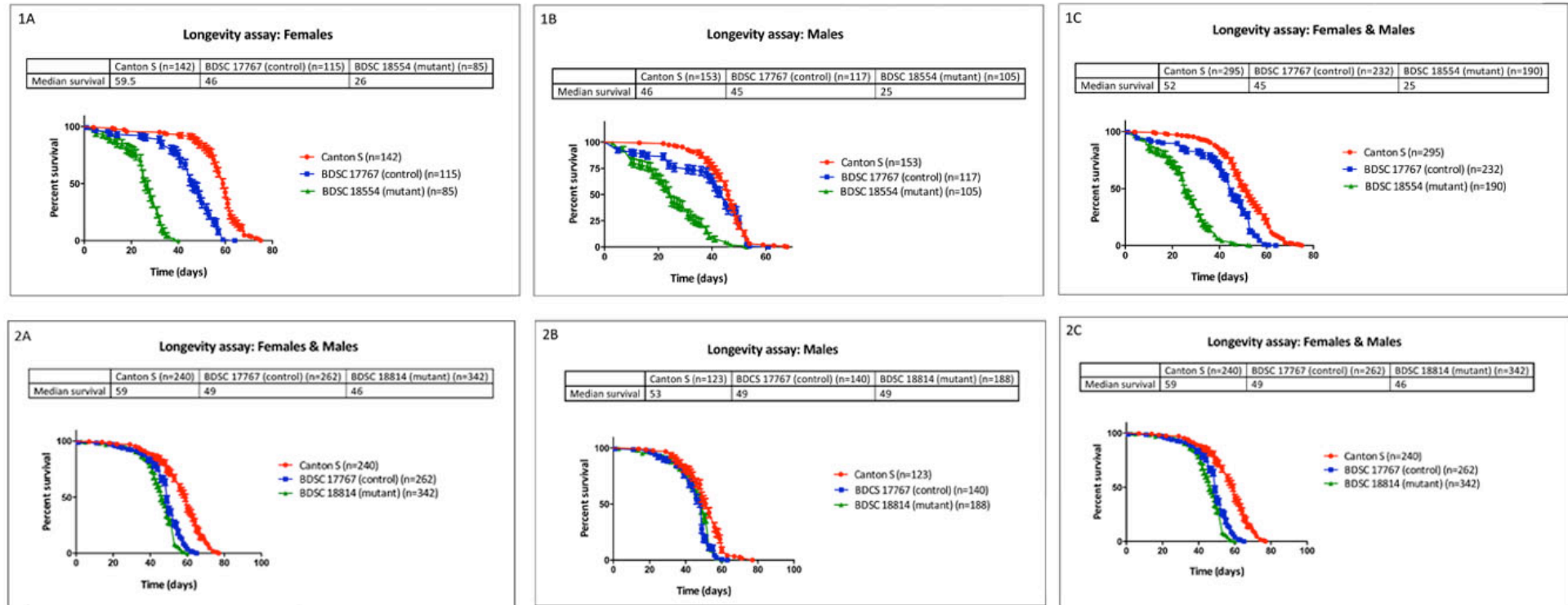


#### 4.2.6 Longevity assay

In order to investigate the effects of *UO* and *CG30016* knockouts on the fly life span, longevity assay was performed. Life span of BDSC 18554 flies is significantly shorter than that of BDSC 17767 control and Canton S flies (Figure 4-10 1A-1C). This is the case for both male and female flies. However, the difference is much greater in females. Median survival values correspond to the life span, and so there is an apparent difference between median survival of mutant flies in comparison to both controls, and the difference is greater in female flies. Interestingly, female Canton S flies live significantly longer than BDSC 17767 flies.

In contrast, there is no difference between life spans of male BDSC 18814, BDSC 17767 control and Canton S flies (Figure 4-10 2B). However, female control and Canton S flies live longer than mutant flies (Figure 4-10 2A). Once again, female Canton S flies also live significantly longer than female BDSC flies. In fact, it seems that the life span of female Canton S flies is much greater than the life span of all other flies tested in the experiment. This suggests that it might be related to adaptive advantage of these flies and does not reflect effects of *CG30016* and *UO* expression/knockout. Altogether, BDSC 18814 flies live shorter than Canton S flies but not BDSC 17767 control flies (Figure 4-10 2C). All life span differences correspond to median survival values. This result is conflicting with previous observations that urate prolongs life span [430] and requires further investigation.

Differences in life span between male and females flies were observed in both mutants. This could be due to differential expression of *CG30016* and *UO* genes in male and female flies. For example, if the expression of these two genes is normally higher in females, the knockout could result in greater changes than in males and affect the life span more. Another possibility is higher sensitivity of female flies to changes in expression of these two genes, having greater impact on their life span.



**Fig. 4-10 Longevity assay, comparison between mutant flies and controls. (1A-1C) Longevity assay in BDSC 18554 flies compared to Canton S wild-type flies and BDSC 17767 controls. Both control and wild-type flies live significantly longer than mutant flies and their mean survival is greater (1C). (1A) Life span of female mutant flies is significantly shorter in comparison to the control and wild-type. This is also true in case of male mutant flies (1B). However, the difference is smaller than in female flies. (2A-2C) Longevity assay in BDSC 18814 flies compared to Canton S wild-type flies and BDSC 17767 controls. (2C) Life span of BDSC 18814 flies is comparable to the life span of BDSC 17767 controls, but shorter than that of Canton S flies. (2A) Female Canton S flies as well as control flies live longer than the mutant flies. (2B) In case of male flies, there is no difference in longevity between the three fly lines.**

## 4.3 Discussion

Our results show that both *UO* and *CG30016* knockouts produce phenotypes in *Drosophila melanogaster*, including MT and abdomen phenotypes, as well as alteration in longevity and survival in oxidative stress conditions. The results indicate that *CG30016* is involved in urate degradation pathway. However, they do not confirm its function as 5-HIU hydrolase.

### 4.3.1 Urate Crystals

Previous research suggested that *UO* starts being expressed in the fruit fly during the L3 larval stage [368, 370]. As a result, urate crystals should be present in the tubules of earlier larvae but not L3 and adult flies. As shown in Chapter 3, we confirmed that *UO* is highly expressed in the L3 stage but not earlier stages. Here, we confirmed this by looking at the presence of urate crystals throughout larval development. Urate crystals are visible in L1, L2 and L3E stages in all tested fly lines, which is consistent with the expression pattern of *UO*. Moreover, in mutant flies, which do not express *UO*, crystals are also present in L3 larvae, confirming the role of *UO* in the breakdown of urate. Interestingly, transparent tubule phenotype was previously described in *rosy* mutants, which are unable to synthesised urate [403]. This is an opposite scenario, where accumulation of urate results in the formation of crystals beyond L3E larval stage.

*CG30016* knockout alone does not seem to affect the formation of urate crystals. In fact, more crystals were observed in *UO* mutant. This is likely due to greater knockout of *UO* in this line as compared to *CG30016* mutant. More crystals are observed in posterior tubules in all fly lines. This could be due to differential expression of *UO* in the MTs due to functional and morphological differences between anterior and posterior tubules [352, 396]. It has been previously suggested that differential orientation of the MTs in the body cavity, corresponds to their function in transport processes from the organs near them (i.e. posterior tubule is involved in transporting metabolites produced by the hindgut). As a result, the posterior tubule is enriched in genes associated with ammonia detoxification of the hindgut [396]. Even though uric acid is not toxic, its accumulation most likely has an impact on the whole insect. Because of this, it would make sense for it to be compartmentalized to the back end of the fly body, away from sensitive tissues. However,

this does not explain why there are also uric acid crystals present in the anterior tubule. In order to understand this, further research is required. It would be interesting to measure the expression of genes related to urate formation and degradation pathways in both anterior and posterior tubules and see whether they have a differential expression depending on tubule orientation.

#### **4.3.2 Inflated tubule and abdomen phenotype and fluid secretion rate**

As expected based on the expression of *UO* and *CG30016* almost exclusively in the MTs, BDSC 18554 mutant flies produced a tubule phenotype, characterised by inflated ureter. It was only present in half of tested flies and was not gender – or tubule orientation – specific. The phenotype was not observed in controls or *UO* mutant. This suggests that the inflation is a result of *CG30016* knockout only, and not accumulation of urate due to *UO* deficiency. The phenotype could be a result of accumulation of a metabolite associated with *CG30016* activity, fluid retention, inflammation, defective osmoregulation or structural abnormality. Mutant flies also appeared to have a bloated abdomen resulting in distorted abdominal patterns. This could also be a result of *CG30016* knockout and requires further examination.

Previous research showed how different genetic manipulations in *Drosophila* can affect fluid secretion rate [71-74]. However, there is no correlation reported between fluid secretion and enzymes/metabolites involved in urate metabolism. Our results showed that *CG30016* knockout did not lead to fluid retention and did not affect basal secretion rate in the fruit fly. This suggests that the inflated ureter and abdomen are a result of either accumulation of a metabolite other than urate, or morphological change in the mutant tissues. Tubule phenotypes including bloated tubule have been previously described [65, 103, 104, 403]. Interestingly, inflated ureter is a novel phenotype that has not been described before. Bloated abdomen has been previously reported as a result of defective osmoregulation in *CIC-a* knockdown flies [69]. In order to test whether the abdominal inflation in BDSC 18554 flies was a result of increased haemolymph water content, further experiments could be performed in the future, including wet-dry weight measurement (as described in [69]). In case of both phenotypes, it would be interesting to characterise them further in order to understand their molecular bases.

### 4.3.3 Is urate an antioxidant or a pro-oxidative factor in the fruit fly?

In agreement with previous studies in *Drosophila melanogaster*, the results suggest that urate acts as antioxidant. This is true for oxidative stress condition induced by H<sub>2</sub>O<sub>2</sub> but not paraquat for experiments performed using *UO* mutant. Interestingly, in flies with both *UO* and *CG30016* knockouts, flies survive better in oxidative stress conditions induced by both H<sub>2</sub>O<sub>2</sub> and paraquat. Moreover, in H<sub>2</sub>O<sub>2</sub> – induced oxidative stress conditions, survival of *CG30016* mutant is greater than that of *UO* mutant. This suggests that the substrate for *CG30016* also acts as an antioxidant in the fly, and in the mutant deficient for both enzymes, urate and another metabolite act synergistically as antioxidants and hence make the flies less sensitive to oxidative stress than urate alone. These results are consistent with a previous study in the fruit fly, which showed that *rosy* mutant, which do not synthesise urate, is hypersensitive to oxidative stress [76]. Altogether, this confirms that urate acts as an antioxidant in oxidative stress conditions.

Our results showed that both mutants had a shorter lifespan than the controls in normal conditions, with exception of *UO* male flies. This could be due to pro-oxidative properties of urate suggested in previous studies. It has been proposed that the main factor involved in lifespan extension and cancer prevention, have been protective mechanisms against ROS [443]. Hence, it would make sense to assume that increased levels of urate, which might induce oxidative stress, has an opposite effect and leads to shorter life span. However, based on our result it is not possible to tell whether shortened lifespan is due to increased oxidative damage or other deleterious effects of increased urate. On the other hand, previous research showed that mutant flies, which are unable to produce urate, have reduced life span [76]. It is hence surprising to see the same effect in mutants unable to degrade urate. This suggests that reduced life span in *CG30016* and *UO* knockout flies might be a result of factors other than urate enrichment. For example, it has been shown that fly's diet can influence life span [444]. Perhaps feeding of these two mutants is altered as a result of the knockout, which in turn affects their longevity. However, in order to test this further research is required.

As was the case with oxidative stress survival, the difference between mutant flies and controls is greater in case of *CG30016* knockout than *UO* knockout. Once again, it suggests that *CG30016* substrate has similar properties to urate, and when both metabolites are enriched, they act synergistically. It would be interesting to investigate

this correlation further, in order to understand its underlying mechanism.

In conclusion, we confirmed here that urate acts as an antioxidant in oxidative stress conditions. We also showed that it reduces life span in normal conditions. However, it is not clear whether it is due to pro-oxidative properties of urate or its other property. Moreover, it is possible that the longevity of mutant flies is not affected by urate accumulation at all. Instead the life span reduction could be caused by another characteristic of the mutant flies. Previous research suggested that antioxidant enzymes improve organism survival in stressful conditions but are not associated with the normal aging process [438]. This contradicts free radical theory of aging but could explain our results. Altogether, based on our research, it is not clear whether urate can act as a pro-oxidative factor and in order to answer this question, further research is required.

## 5 Metabolomics analysis in *Drosophila melanogaster*

### 5.1 Introduction

Metabolomics has been previously used to study organismal function in *Drosophila melanogaster*. The method has been applied to fly research in functional genomics [62, 106, 177, 187], development [445-447], effects of different diets and dietary restrictions on the metabolome [175, 448-450], *Drosophila* Parkinson's disease model [451, 452], metabolic responses to different environmental conditions and stresses (such as light and temperature) [173, 175, 453-455], as well as pesticides [90]. Metabolomics provided metabolic profiles of flies responding to different stimuli, including environmental changes and genetic perturbation. This aided the understanding of molecular basis of different adaptive mechanisms, drug/pesticide action, gene function and many more. For example, metabolomics analysis has been applied to study how insecticide permethrin affects metabolic pathways. This revealed that lipid and energy metabolism were altered upon application of the drug [90]. Another study showed that *Drosophila* models of Parkinson's disease (PD) undergo similar metabolic alterations to PD patients. As a result, it validated the fly as a model for studying neurodegenerative properties of environmental chemicals and molecular basis of PD [451].

Most of these studies focused on the whole fly rather than individual fly tissues. However, it is clear by looking at FlyAtlas data, that single tissue metabolomes need to be investigated. FlyAtlas is a tissue-specific transcriptomic atlas for *Drosophila melanogaster*, which provides information on gene enrichment in different adult and larval tissues [6]. FlyAtlas revealed that different tissues differ from each other and the whole fly in terms of gene expression. Moreover, it is estimated that each tissue contributes around 5% of the whole-organism transcriptome, so even large changes in gene expression in one tissue would be underrepresented in the whole fly transcriptome [6]. Hence, we are expecting to see a similar pattern when looking at metabolomes of different fly tissues; namely that the whole organism metabolome would be a poor predictor of what was taking place in individual tissues. Interestingly, previous research in other species has showed that metabolite profiles are highly tissue-specific. For example, a study in

*Arabidopsis*, has revealed that various phytochemicals were produced by the plant in a tissue-specific way [125]. The data were also correlated to *Arabidopsis* transcriptome, which showed that accumulation of metabolites in a given tissue corresponded to enrichment of a gene responsible for its biosynthesis in the same tissue [138].

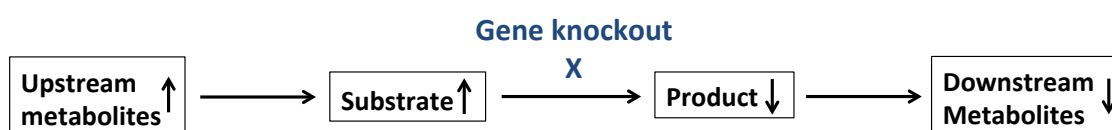
Several studies in *Drosophila* have looked at the correlation of the metabolomics with other omics approaches, including genomics and transcriptomics [447, 450, 456]. For example, Hoffman *et al.* showed that there is an association between metabolomics and age, sex and genotype in predicting life expectancy. The study identified metabolic pathways that helped to bridge the gap between the genome and life expectancy [456]. Another study looked at the correlation between transcriptome, metabolome and phenotypic manifestations linked to human metabolic syndrome (body weight, total sugar, and total triglycerides) of flies raised on different diets. It revealed that there is little association between transcriptome/metabolome and the gross phenotype of the flies, which depends more on the type of diet consumed. However, few genes and metabolites were correlated to metabolic syndrome-like phenotypes in all flies on different diets. For example, metabolites linked to body weight and total triglycerides and sugars were L-DOPA (L-3,4-dihydroxyphenylalanine), fatty acids, and BCAAs (branch chain amino acids), and correlated genes were related to ATP synthesis, TCA cycle and pupariation [450].

Finally, several studies have looked at metabolite levels of individual *Drosophila* tissues, including lipid composition of fat body, wing disc, salivary gland, gut and brain of fruit fly larvae [457]. This study revealed that there were significant differences in the lipid content between different tissues. Another study investigated differences in metabolism between head, thorax and abdomen [396]. Moreover, Laye *et al.* studied dietary restriction (DR) and its effect on age-related diseases correlated with the metabolome of the whole fly, head, thorax, and abdomen. The study revealed that DR had a great effect on the metabolomes of different tissues, and particularly on the metabolic pathways related to aging. Moreover, it showed that some metabolites enriched in DR were common to all tissues. However, each tissue also had a unique metabolic profile as a result of DR [449].



### 5.1.1 How do single-gene mutations affect the fly metabolome?

Previous studies on several mutations in the fruit fly genome (including *ry*, *cho*, *y* and *mal*) investigated their effect on the metabolome [62, 106, 177, 187]. This research revealed that by using LC-MS-based metabolomics, it is possible to observe effects of single-gene mutations on the pathway the gene is involved in, and also on other apparently unrelated pathways, and the global metabolome. The general principle is that if an enzyme is knocked down we expect to observe increase in concentration of upstream metabolites and decrease in downstream metabolites (Figure 5-1).



**Fig. 5-1 The principle of single-gene knockout metabolomics. When an enzyme-encoding gene is knocked out, metabolite levels upstream from the lesion increase, and the ones downstream decrease.**

### 5.1.2 Metabolomics approaches to studying urate degradation mutants

As shown in previous chapter, phenotypic analysis of mutant flies did not establish the function of *CG30016* gene in *Drosophila*. However, the results suggest that urate degradation pathway is disrupted in the knockout flies due to accumulation of urate in the tubules. If the hypothesis that *CG30016* is 5-HIU hydrolase is correct, we should observe increased levels of both urate and 5-HIU in the mutants as well as decreased levels of OHCU and allantoin. Additionally, by observing how the mutation affects global metabolome of the fly, we could gain an understanding of how the disruption of urate degradation pathway leads to observed phenotypes.

In this chapter we summaries the results obtained from studying metabolomes of urate degradation mutants using LC-MS. We looked at how single-gene knockouts affect whole fly and tubule metabolomes. Moreover, metabolomes of the whole fly as well as three different tissues were generated. Finally, we validated the use of this metabolomics platform for the study of fly secretion fluid.

## 5.2 Results

### 5.2.1 Tissue-specific metabolomes of *Drosophila melanogaster*

Data presented in this section were obtained and analysed in collaboration with V. Chintapali and D. Watson and published [186]. Two different analyses were performed. The experiments performed at Strathclyde University (SU) compared whole fly, head, crop, mid gut, anterior tubules, posterior tubules, hind gut, ovaries, testes, accessory glands and cuticle. The analysis carried out at Polyomics Facility (PF), compared whole fly, central nervous system, Malpighian tubules and the gut. For the purpose of this thesis, we focused on selected tissues and presented a summary below. In case of PF data, all data collection, processing and analysis were performed by myself, and they are presented in this thesis. The data obtained from SU were collected and processed by our collaborators. Data analysis and presentation was discussed and performed by the collaborators and myself. SU data are presented here for the purpose of comparison of the two sets of results, and a validation of LC-MS metabolomics as a tool for obtaining tissue-specific metabolomes.

SU identified 242 metabolites that were run alongside 82 standards. Metabolites were identified and quantified using Xcalibur software combined with Sieve Software 1.3 (Thermo Fisher Co.). Except metabolites that matched masses and retention times of standards, all metabolites were putatively identified. All masses were within 1.5 ppm of the exact molecular formula. This means that the only competing metabolites within the database search were isomers. SU results are represented as heat maps of peak areas normalised to the protein content of each tissue. Repeatability of this method was also expressed as RSD values for the peak areas across four replicates. In some instances these values were very low, which reflected variability between individual samples. Confidence levels of each identified metabolite were also assigned: 1 – exact mass and RT matched a standard, 2 – accurate mass matched, isomers possible.

Measurements obtained from PF identified 202 metabolites and run 134 standards (Appendix V and VI). The instrument settings as well as data analysis are described in Materials and Methods (Chapter 2.16). For each metabolite, mean signal intensity was calculated in each sample (based on four biological replicates). IDEOM algorithm was used to express relative abundance of each metabolite. Relative abundance of each

metabolite in the control (whole fly (WF)) was expressed as '0' (metabolite absent in the sample) or '1' (metabolite present in the sample). In all other samples relative abundance was expressed as a fold change relative to the control group calculated from mean signal intensity of each metabolite versus mean signal intensity of this metabolite in the control. Metabolite levels were coloured according to relative abundance (blue = low, red = high). Moreover, t-test was performed for each metabolite as part of IDEOM pipeline, and where the P-value (each sample t-tested against control group) was less than 0.05, the data was analysed further.

In case of tissue-specific metabolomes, the most enriched metabolites in individual tissues were calculated using an additional formula:

Tissue enrichment = Relative abundance in a tissue of interest/(the sum of abundance in all other tissues+1)

Both analyses are semi-quantitative and hence represent the discovery stage and generate hypotheses. Since FlyAtlas revealed differences in gene expression between different tissues and the whole fly we investigate here whether the same is true for tissue metabolomes. We expect that any differences between tissue metabolomes are likely to have functional significance. As was the case with the fly transcriptome [6], there are clear differences between tissue metabolomes as well as some similarities.

#### **5.2.1.1 Tissue metabolomes differ from each other and from the whole fly**

Following data processing by IDEOM, additional multivariate and univariate statistical analyses were performed using MetaboAnalyst (<http://www.metaboanalyst.ca>), an open-source platform for metabolomics data analysis [258, 458]. In order to group samples according to similar abundance profiles, hierarchical clustering analysis, dendrogram, was performed (dissimilarity measure used was Spearman's rank correlation, and clustering method was Ward's linkage). Moreover, PCA was carried out in order to determine how varied the replicates were within each sample, whether they clustered together, and whether there were differences between metabolomes of analysed tissues. Finally, one-way ANOVA was performed, in order to establish whether there were significant differences between samples, i.e. whether the metabolomes of different tissues varied significantly.

Our results clearly showed that tissue metabolomes differ significantly from each other

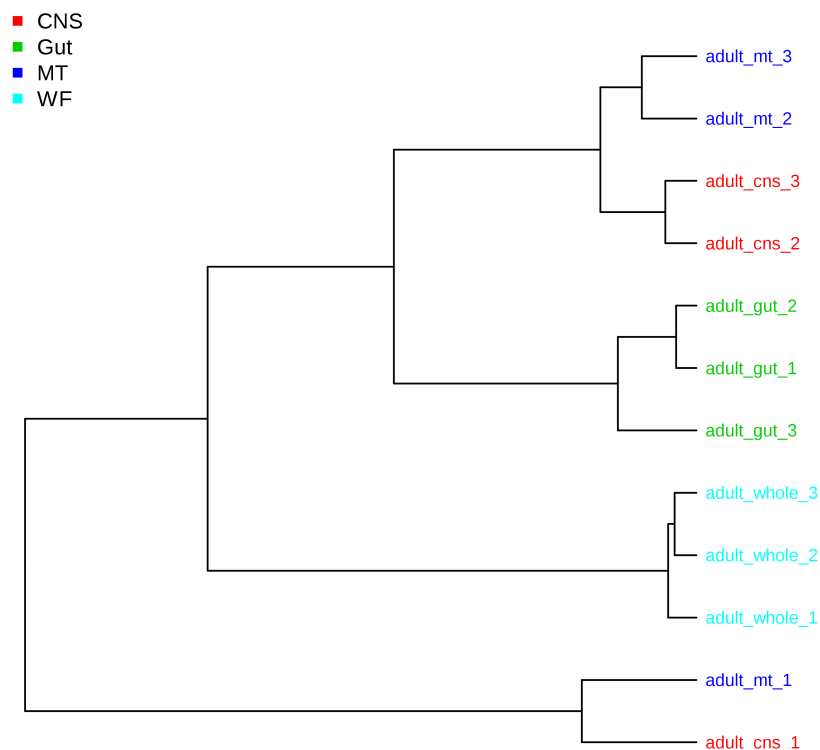
and from the whole fly metabolome (Figures 5-2 and 5-3). Hierarchical clustering analysis revealed that generally samples from the same tissues clustered together and were hence related to each other but not other tissues (Figure 5-2A). The whole fly as well as the gut samples clustered together and separately from other tissues, and so did two out of three replicates of CNS and MTs samples. On the other hand, cns1 and mt1 samples were more related to each other than to the other two replicates of the same tissue. This confirms that tissue metabolomes differ from each other but also shows that there is variability between replicates of the same sample.

Similar results were illustrated by PCA (Figure 5-2B), where whole fly replicates clustered together and separately from other samples, and there was variability between replicates of different tissue samples. However, PCA showed that the biggest variance was present in cns1 sample, which did not cluster with other CNS replicates.

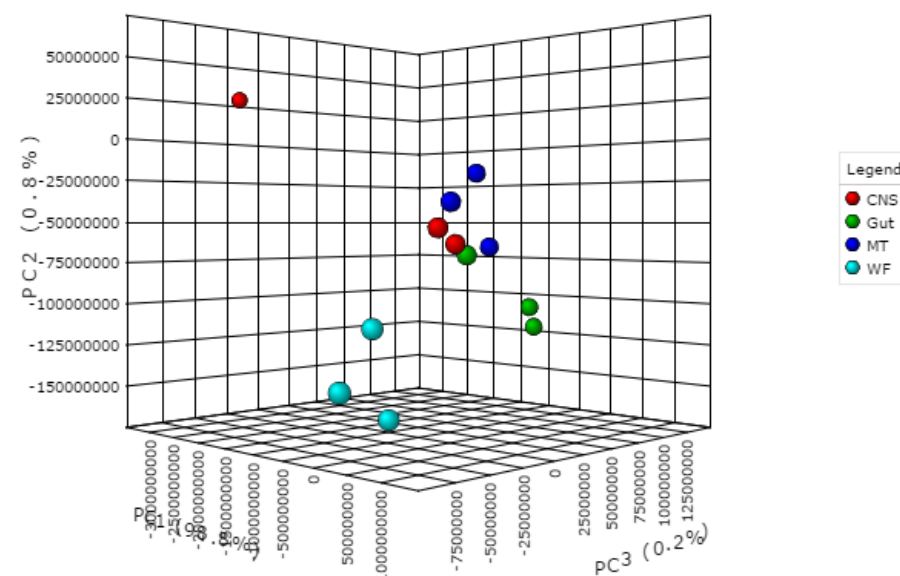
Finally, one-way ANOVA results are illustrated in Figure 5-3. Each dot represents a single metabolite; red dots represent metabolites, for which the abundance was significantly different between the samples, green dots represent metabolites, for which the abundance was not significantly different between the tissues and the whole fly. These results confirmed that whereas some metabolite abundances differ between tissues, other metabolites had consistent abundance across samples.

Altogether, these results showed that tissue metabolomes differed from each other and from the whole fly metabolome. However, some metabolites are enriched across all the tissues, which shows that some metabolites and metabolic pathways are conserved in the whole fly and in all tested tissues. This might reflect that whereas some organismal functions are tissue-specific, others are performed by multiple tissues. Finally, presented analyses revealed that tissue metabolomes vary between replicates. This showed the importance of standardisation of data collection protocol, as well as using as many replicates as is feasible in order to obtain the most real picture of tissue metabolome.

A



B



**Fig. 5-2 Hierarchical clustering analysis (A) and PCA (B) of tissue metabolomes. (A) Dendrogram illustrates clustering of all samples, showing that all replicates of the whole fly and gut samples cluster together, whereas in case of MTs and CNS samples, two out of three replicates cluster together. On the other hand, one CNS (cns1) and one MTs (mt1) sample cluster with each other rather than replicates of the same tissue. (B) PCA analysis illustrates that whole fly (blue), MTs (navy blue) and gut samples (green) cluster together, whereas one CNS replicate (cns1) differs from the other two CNS samples (red). Generated using MetaboAnalyst (<http://www.metaboanalyst.ca>).**

## One-way ANOVA

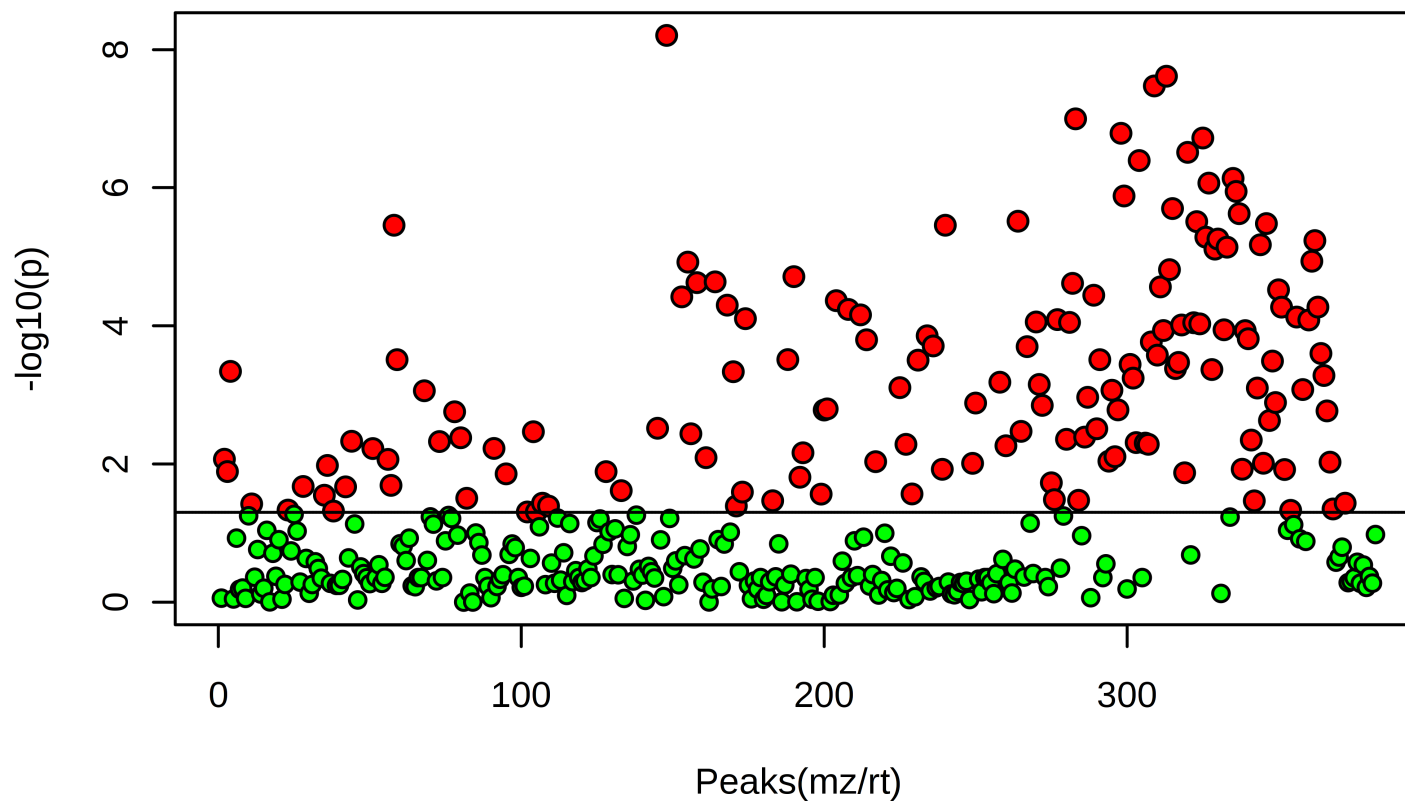


Fig. 5-3 Results of one-way ANOVA. Each dot represents a single metabolite; red dots represent metabolites, for which the abundance was significantly different between the samples, green dots represent metabolites, for which the abundance was not significantly different between the samples. Peaks (mz/rt) refer to detected metabolites with a unique mass-to-charge ratio and a unique retention time. Generated using MetaboAnalyst (<http://www.metaboanalyst.ca>).

### 5.2.1.2 Most abundant metabolites – amino acids

Amino acids were among the most abundant metabolites across all samples. Figure 5-4 shows the top 25 most abundant metabolites in the whole fly and averaged over all tissues. The single most abundant metabolite in the whole fly is proline, which is also highly abundant across all the tissues. Other metabolites including histidine, alanine, glutamine, arginine etc. are also highly abundant. However, it is interesting to see that except proline, other amino acids have a tissue-specific pattern of enrichment. Moreover, the levels of amino acids are lower in Malpighian tubules in comparison to other tissues. This could reflect the role of MTs of generating the primary urine. This involves selective reabsorption of desirable metabolites, such as amino acids. Hence, low abundance of amino acids in MTs could reflect their role in amino acid recycling back to the haemolymph.

M/Z	RT	Name	Level	Iso	RSD %	WF	HD	Crop	MG	AT	PT	HG	OV	TEST	ACCG	CUT
116.0705	15.1	D-Proline	1	0	14.8											
156.0767	24.0	L-Histidine	1	0	33.2											
90.05495	16.3	Alanine	1	2	9.4											
147.0763	16.8	glutamine	1	5	17.1											
90.05491	17.5	beta alanine	1	2	6.5											
175.119	25.8	L-Arginine	1	0	11.4											
148.0603	16.0	glutamate	1	3	19.8											
166.0532	16.6	Methionine sulfoxide	2	0	31.8											
118.0862	15.3	betaine	1	2	17.3											
184.0734	22.1	Choline phosphate+	2	0	15.4											
162.1124	16.5	Carnitine	1	0	14.4											
132.1018	12.9	Leucine/isoleucine	1	4	9.4											
225.0869	14.0	3-Hydroxy-L-kynurenine	2	1	21.8											
118.0862	14.4	valine	1	2	22.4											
104.107	17.4	Choline	1	0	18.2											
204.123	13.8	L-Acetylcarnitine	1	0	13.2											
154.0974	15.9	N-acetylhistamine	2	1	21.4											
206.0448	8.5	Hydroxykynurenate	2	0	6.7											
209.092	12.6	kynurenine	1	0	38.7											
120.0655	16.4	Threonine	1	1	9.1											
106.0498	17.2	Serine	1	0	11.4											
166.0862	12.3	phenylalanine	1	0	15.0											
126.0219	15.4	Taurine	1	0	8.1											
348.0703	16.5	Adenosine monophosphate	1	1	9.0											
104.0706	16.8	aminobutyric acid	2	3	4.9											

**Fig. 5-4 The top 25 most abundant metabolites by area response in the whole fly and ten fly tissues. Red = area > 10<sup>8</sup> Yellow = area > 10<sup>7</sup> Blue = area > 10<sup>5</sup>; WF - whole fly, HD – head, MG – midgut, AT – anterior tubule, PT- posterior tubule, HG – hindgut, OV – ovaries, TEST – testes, ACCG – accessory glands, CUT – cuticle, from [186].**

In agreement with SU data, the most enriched metabolites in the dataset analysed by PF are part of amino acid metabolism (Table 5-1). Moreover, other highly enriched metabolites belong to carbohydrate metabolism, and metabolism of vitamins and cofactors. Moreover, some metabolites have not been assigned to any known metabolic pathway. This illustrates the limitations of the software used and/or current knowledge gaps. Once again, it is clear by looking at these results that tissue metabolomes differ

from each other and from the whole fly. However, some metabolites are enriched across all the tissues.



**Table 5-1 The most enriched metabolites across all tissues. Results are represented as a ratio of relative abundance in a given tissue in comparison to other samples, with the whole fly (WF) as a control. Metabolite levels are coloured according to relative intensity (blue = low, red = high, blue<yellow<orange<red).**

Mass	RT	Formula	Putative metabolite	Map	CNS	Gut	MTs	WF
195.09	5.79	C <sub>10</sub> H <sub>13</sub> NO <sub>3</sub>	Damascenine	No known pathway	148.92	106.28	70.73	0.00
118.03	6.84	C <sub>4</sub> H <sub>6</sub> O <sub>4</sub>	Succinate	Carbohydrate Metabolism	4.79	63.15	34.16	1.00
191.06	6.20	C <sub>7</sub> H <sub>13</sub> NO <sub>3</sub> S	N-Acetylmethionine	No known pathway	7.49	33.03	8.56	0.00
123.03	8.13	C <sub>6</sub> H <sub>5</sub> NO <sub>2</sub>	Nicotinate	Metabolism of Cofactors and Vitamins	10.49	30.04	7.65	1.00
129.04	7.64	C <sub>5</sub> H <sub>7</sub> NO <sub>3</sub>	L-1-Pyrroline-3-hydroxy-5-carboxylate	Amino Acid Metabolism	2.27	7.34	33.99	1.00
175.06	6.02	C <sub>10</sub> H <sub>9</sub> NO <sub>2</sub>	Indole-3-acetate	Amino Acid Metabolism	35.87	7.25	15.23	0.00
241.11	7.24	C <sub>25</sub> H <sub>30</sub> N <sub>4</sub> O <sub>6</sub>	Phe-Gly-Pro-Tyr	Peptide(tetra-)	3.12	15.15	14.45	1.00
148.05	5.79	C <sub>9</sub> H <sub>8</sub> O <sub>2</sub>	trans-Cinnamate	Amino Acid Metabolism	21.67	16.07	12.99	0.00
187.08	5.80	C <sub>8</sub> H <sub>13</sub> NO <sub>4</sub>	6-Acetamido-2-oxohexanoate	Amino Acid Metabolism	9.20	12.69	9.64	1.00
758.51	4.78	C <sub>48</sub> H <sub>70</sub> O <sub>7</sub>	(3R,2'S)-Myxol 2'-(2,4-di-O-methyl- $\alpha$ -L-fucoside)	Biosynthesis of Secondary Metabolites	4.65	2.05	6.54	0.00
163.10	5.88	C <sub>10</sub> H <sub>13</sub> NO	N-Acetylphenylethylamine	No known pathway	20.05	13.74	6.53	1.00

175.03	5.86	C9H19O12P	nonulose 9-phosphate	No known pathway	10.67	8.88	6.06	1.00
190.08	5.81	C8H14O5	(R)-3-((R)-3-Hydroxybutanoyloxy)butanoate	Carbohydrate Metabolism	8.71	5.32	5.45	1.00
173.07	6.06	C7H11NO4	N-Acetyl-L-glutamate 5-semialdehyde	Amino Acid Metabolism	4.64	9.41	4.93	1.00
145.07	6.89	C6H11NO3	4-Acetamidobutanoate	Amino Acid Metabolism	1.65	10.36	4.37	1.00
141.04	6.20	C6H7NO3	6-oxo-1,4,5,6-tetrahydronicotinate	Metabolism of Cofactors and Vitamins	7.10	9.04	4.17	1.00
138.04	12.81	C6H6N2O2	Urocanate	Amino Acid Metabolism	1.98	6.88	3.92	1.00
159.09	6.08	C7H13NO3	5-Acetamidopentanoate	Amino Acid Metabolism	2.22	4.06	3.91	1.00
113.05	6.08	C5H7NO2	1-Pyrroline-2-carboxylate	Amino Acid Metabolism	3.68	7.14	3.65	1.00

### 5.2.1.3 Metabolomes of Malpighian tubules

SU results revealed that in both posterior and anterior tubules, the most abundant metabolites are part of tryptophan metabolism, including kynurenine, tryptophan, hydroxykynurenate etc. (Figure 5-5). This is particularly interesting considering that MT metabolome is depleted for most amino acids (Figure 5-5). Tryptophan is a precursor of a visual pigment, and it is known that MTs store visual pigment precursors [459]. Moreover, previous research revealed that tryptophan is actively transported into the tubule [460]. Hence, it makes sense the tubule is enriched in tryptophan and related metabolites.

M/Z	RT	Name	Level	Iso	RSD %	WF	HD	Crop	MG	AT	PT	HG	OV	TEST	ACCG	CUT
225.0869	14.0	3-Hydroxy-L-kynurenine	2	0	24.8	Red	Red	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow
209.092	12.6	kynurenine	1	0	14.1	Red	Red	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow
206.0448	8.5	6-Hydroxykynurenate	2	0	31.2	Red	Red	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow
205.0972	13.0	Tryptophan	1	0	34.3	Red	Red	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow
237.0869	14.6	N-Formylkynurenine	2	0	71.1	Red	Red	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow
221.092	13.6	Hydroxytryptophan	2	1	39.1	Red	Red	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow
243.0876	6.1	Lumichrome	2	0	34.0	Red	Red	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow
377.1454	8.6	Riboflavin	1	0	55.8	Red	Red	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow
298.1144	10.3	Methylguanosine	2	3	109.8	Red	Red	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow
180.088	12.9	Dimethylguanine	2	2	49.8	Red	Red	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow
136.0618	14.8	Adenine	1	0	41.4	Red	Red	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow
174.0551	12.6	Quinaldic acid	2	0	50.4	Red	Red	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow

**Fig. 5-5 The most abundant metabolites by area response in Malpighian tubules. Red = area > 10<sup>8</sup> Yellow = area > 10<sup>7</sup> Blue = area > 10<sup>5</sup>; WF - whole fly, HD – head, MG – midgut, AT – anterior tubule, PT- posterior tubule, HG – hindgut, OV – ovaries, TEST – testes, ACCG – accessory glands, CUT – cuticle, from [186].**

The analysis performed at PF showed similar results. Among the most abundant metabolites in the tubule were L-kynurenine and L-Formylkynurenine, which are part of tryptophan metabolism (Table 5-2). This validates the results obtained at SU. Moreover, metabolites involved in purine metabolism were detected in the tubule, including xanthine, hypoxanthine, inosine, deoxyinosine, deoxyadenosine, adenosine and adenine (Appendix VI). This confirmed that the analysis of urate degradation mutants could be performed using MTs rather than the whole fly.

**Table 5-2 Metabolites enriched in Malpighian tubules (the highest tissue enrichment value for MTs). Results are represented as a ratio of relative abundance in a given tissue in comparison to other samples. Metabolite levels are coloured according to relative intensity (blue = low, red = high, blue<yellow<orange<red).**

Mass	RT	Formula	Putative metabolite	Pathway	CNS	Gut	MTs	WF
129.04	7.64	C <sub>5</sub> H <sub>7</sub> NO <sub>3</sub>	L-1-Pyrroline-3-hydroxy-5-carboxylate	Arginine and proline metabolism	2.27	7.34	33.99	1.00
165.05	17.63	C <sub>5</sub> H <sub>11</sub> NO <sub>3</sub> S	L-Methionine S-oxide	Methionine metabolism	0.00	0.27	16.81	1.00
236.08	13.17	C <sub>11</sub> H <sub>12</sub> N <sub>2</sub> O <sub>4</sub>	L-Formylkynurenine	Tryptophan metabolism	0.00	0.30	5.48	1.00
109.05	15.31	C <sub>6</sub> H <sub>7</sub> NO	2-Aminophenol	Tryptophan metabolism	0.62	0.18	5.28	1.00
208.08	13.46	C <sub>10</sub> H <sub>12</sub> N <sub>2</sub> O <sub>3</sub>	L-Kynurenine	Tryptophan metabolism	0.00	0.03	5.27	1.00
191.04	8.87	C <sub>7</sub> H <sub>5</sub> N <sub>5</sub> O <sub>2</sub>	6-formyl-H <sub>2</sub> -pterin	No known pathway	0.04	0.13	4.59	1.00
173.14	14.98	C <sub>9</sub> H <sub>19</sub> NO <sub>2</sub>	Muscarine	No known pathway	0.00	1.41	3.76	1.00
376.14	8.00	C <sub>17</sub> H <sub>20</sub> N <sub>4</sub> O <sub>6</sub>	Riboflavin	Riboflavin metabolism	0.00	0.13	2.91	1.00
154.03	6.06	C <sub>7</sub> H <sub>6</sub> O <sub>4</sub>	2,5-Dihydroxybenzoate	Tyrosine metabolism	0.00	0.00	1.85	0.00
464.46	5.25	C <sub>31</sub> H <sub>60</sub> O <sub>2</sub>	Hentriacontane-14,16-dione	No known pathway	0.00	0.00	1.47	0.00

### 5.2.1.4 Metabolomes of the gut

The midgut is mostly responsible for digestion and nutrient absorption. Among the most abundant metabolites in the midgut are acylcarnitines (Figure 5-6). Acylcarnitines are responsible for buffering the levels of CoA within the mitochondria as well as transport of fatty acids. FlyAtlas revealed that most dietary lipases are highly expressed in the midgut [4]. This is consistent with the fact that the midgut is the major site of the fatty acid uptake and processing.

M/Z	RT	Name	Level	Iso	RSD	WF	HD	Crop	MG	AT	PT	HG	OV	TEST	ACCG	CUT
330.2272	6.0	Ketodecanoylcarnitine	2	0	29.7											
308.0909	14.9	Glutathione	1	0	24.1											
412.3054	5.9	Hydroxyhexadecadienoylcarnitine	2	0	24.9											
384.2742	5.9	Hydroxytetradecadienoylcarnitine	2	0	20.8											
386.2898	5.9	Hydroxytetradecenoylcarnitine	2	0	58.8											
344.2787	10.7	Dodecanoylcarnitine	2	0	47.5											
370.2947	10.6	Tetradecenoylcarnitine	2	0	54.1											
189.1235	16.2	N-Alpha-acetyllysine	1	1	32.1											
251.0695	14.4	Glutamylcysteine	2	0	19.6											
316.2481	10.9	Decanoylcarnitine	2	0	30.7											
229.1546	14.8	Leucyl-proline	2	1	38.5											
342.263	10.7	Dodecenoylcarnitine	2	0	28.6											
312.2168	6.0	Decadienoylcarnitine	2	0	25.6											
288.2167	6.0	Octanoylcarnitine	2	0	52.0											
388.3054	10.8	Hydroxymyristoylcarnitine	2	0	59.9											
384.1148	9.7	Succinyladenosine	2	0	91.9											
288.2168	11.2	Octanoylcarnitine	2	0	20.6											
198.0874	17.1	N-Acetyl-L-histidine isomer	2	1	83.1											
368.2787	10.6	Tetradecadienoylcarnitine	2	0	110.3											
118.1226	16.1	Methylcholine	2	0	76.1											
260.1856	11.6	Hexanoylcarnitine	2	0	15.2											
360.2742	11.1	Hydroxylauoylcarnitine	2	0	59.0											
302.2325	6.0	Nonanoylcarnitine	2	0	63.3											
442.3515	10.6	Hydroxyoctadecenoylcarnitine	2	0	54.7											
256.0815	14.4	Nicotinate D-ribonucleoside	2	0	53.9											
396.3106	10.6	Hexadecadienoylcarnitine	2	0	133.9											
314.2325	6.0	Decenoylcarnitine	2	0	25.2											
268.0849	16.8	S-Ribosylhomocysteine	2	0	106.7											

**Fig. 5-6 The most abundant metabolites in fly midgut. Red = area > 10<sup>8</sup> Yellow = area > 10<sup>7</sup> Blue = are > 10<sup>5</sup>; WF - whole fly, HD – head, MG – midgut, AT – anterior tubule, PT- posterior tubule, HG – hindgut, OV – ovaries, TEST – testes, ACCG – accessory glands, CUT – cuticle, from [186].**

The results obtained from PF differ from the ones above. This might be because the analysis was performed on the whole gut tissue rather than the midgut alone. The results showed that the most abundant metabolites in the whole gut are part of amino acid metabolism, including phenylalanine metabolism (Table 5-3). Moreover, the tissue is rich in peptides. This corresponds to the role of the gut in digestion, and might reflect the results of protein degradation as well as abundance of digestive enzymes and peptides in the tissue.

**Table 5-3 Metabolites enriched in the gut (the highest tissue enrichment value in the gut). Results are represented as a ratio of relative abundance in a given tissue in comparison to other samples Metabolite levels are coloured according to relative intensity (blue = low, red = high, blue<yellow<orange<red).**

Mass	RT	Formula	Putative metabolite	Pathway	CNS	Gut	MTs	WF
207.09	5.94	C <sub>11</sub> H <sub>13</sub> NO <sub>3</sub>	N-Acetyl-L-phenylalanine	Phenylalanine metabolism	0.00	27.14	0.00	0.00
219.11	6.76	C <sub>9</sub> H <sub>17</sub> NO <sub>5</sub>	Pantothenate	beta-Alanine metabolism	2.17	17.72	0.98	1.00
191.06	6.20	C <sub>7</sub> H <sub>13</sub> NO <sub>3</sub> S	N-Acetylmethionine	No known pathway	7.49	33.03	8.56	0.00
118.03	6.84	C <sub>4</sub> H <sub>6</sub> O <sub>4</sub>	Succinate	Phenylalanine metabolism	4.79	63.15	34.16	1.00
169.07	16.69	C <sub>8</sub> H <sub>11</sub> NO <sub>3</sub>	Pyridoxine	Vitamin B6 metabolism	0.00	16.40	8.65	1.00
123.03	8.13	C <sub>6</sub> H <sub>5</sub> NO <sub>2</sub>	Nicotinate	Nicotinate and nicotinamide metabolism	10.49	30.04	7.65	1.00
146.06	6.30	C <sub>6</sub> H <sub>10</sub> O <sub>4</sub>	(S)-2-Aceto-2-hydroxybutanoate	Valine, leucine and isoleucine biosynthesis	2.32	8.71	1.88	1.00
145.07	6.89	C <sub>6</sub> H <sub>11</sub> NO <sub>3</sub>	4-Acetamidobutanoate	Arginine and proline metabolism	1.65	10.36	4.37	1.00
582.44	5.77	C <sub>41</sub> H <sub>58</sub> O <sub>2</sub>	Spheroidenone	Carotenoid biosynthesis	0.00	2.58	0.06	1.00
114.03	6.35	C <sub>5</sub> H <sub>6</sub> O <sub>3</sub>	2-Hydroxy-2,4-pentadienoate	Phenylalanine metabolism	0.00	20.85	18.99	0.00
400.23	11.89	C <sub>18</sub> H <sub>32</sub> N <sub>4</sub> O <sub>6</sub>	Ala-Leu-Thr-Pro	Hydrophobic peptide	0.00	1.98	0.00	1.00
138.04	12.81	C <sub>6</sub> H <sub>6</sub> N <sub>2</sub> O <sub>2</sub>	Urocanate	Histidine metabolism	1.98	6.88	3.92	1.00
473.21	11.89	C <sub>19</sub> H <sub>31</sub> N <sub>5</sub> O <sub>9</sub>	Glu-Thr-Gln-Pro	Acidic peptide	0.00	1.70	0.00	1.00

173.07	6.06	C7H11NO4	N-Acetyl-L-glutamate 5-semialdehyde	Arginine and proline metabolism	4.64	9.41	4.93	1.00
426.25	11.76	C20H34N4O6	Ile-Thr-Pro-Pro	Hydrophobic peptide	0.00	1.62	0.00	1.00
241.11	7.24	C25H30N4O6	Phe-Gly-Pro-Tyr	Hydrophobic peptide	3.12	15.15	14.45	1.00
113.05	6.08	C5H7NO2	1-Pyrroline-2-carboxylate	Arginine and proline metabolism	3.68	7.14	3.65	1.00
244.07	9.25	C9H12N2O6	Uridine	Pyrimidine metabolism	0.11	1.86	0.51	1.00

### 5.2.1.5 Metabolome of *Drosophila melanogaster* head and central nervous system

*Drosophila* head contains primarily the head capsule, some fat body, brain and compound eyes. As a result, the most abundant metabolites in the head are xanthommatin and red pterin metabolites (Figure 5-7). This is consistent with the fact that the head is enriched in pigment-processing genes, including *sepia*, *Plum*, *Henna*, *vermillion* and *brown* (FlyAtlas.org).

M/Z	RT	Name	Level	Iso	RSD	WF	HD	Crop	MG	AT	PT	HG	OV	TEST	ACCG	CUT
206.0448	8.5	6-Hydroxykynurenate	2	0	26.7											
369.1528	17.9	Drosoplerin	2	0	18.5											
184.0734	22.1	Choline phosphate+	2	0	6.2											
162.055	9.1	Indole carboxylic acid	2	1	14.2											
238.0934	12.0	Sepiapterin	1	2	46.5											
424.0773	11.7	Xanthommatin	2	0	31.9											
238.0933	9.4	Biopterin isomer	2	2	29.4											
164.0567	11.3	Pterin	2	1	27.4											
176.0553	8.5	N-Formylglutamic acid	2	1	23.1											
240.109	12.9	Dihydrobiopterin isomer	2	3	37.2											
196.0829	11.1	Hydroxymethyldihydropterin	2	0	21.6											
130.0974	15.5	guanidino butanal	2	0	42.0											
240.109	14.5	Dihydrobiopterin isomer	2	3	36.6											
268.1039	11.5	*Deoxyguanosine	1	1	30.6											
236.0777	9.2	Dehydrosepiapterin	2	0	32.4											
137.0458	10.0	Hypoxanthine	1	0	14.0											
182.0673	11.2	Dihydroxanthopterine	2	0	42.1											
196.083	10.0	Hydroxymethyldihydropterin	2	0	20.4											
184.083	17.2	Tetrahydroxanthopterine	2	0	21.0											
180.0517	12.1	Xanthopterine	2	1	15.5											
364.0649	16.5	GMP	1	0	25.0											
298.1144	10.3	**Methylguanosine	1	0	29.1											
254.0883	10.5	Hydroxybiopterin	2	2	41.8											
197.067	9.2	Dimethyluric acid	2	2	29.3											
180.088	11.6	Dimethylguanine	2	2	23.6											
171.0514	9.3	5-Ureido-4-imidazole carboxylate	2	0	12.6											
253.0931	10.0	***Deoxyinosine	1	0	13.9											
349.0541	14.8	Inosine monophosphate	1	1	16.2											
111.0553	17.1	Imidazole-4-acetaldehyde	2	0	18.2											

**Fig. 5-7 The most abundant metabolites in *Drosophila melanogaster* head. Red = area > 10<sup>8</sup> Yellow = area > 10<sup>7</sup> Blue = area > 10<sup>5</sup>; WF - whole fly, HD – head, MG – midgut, AT – anterior tubule, PT- posterior tubule, HG – hindgut, OV – ovaries, TEST – testes, ACCG – accessory glands, CUT – cuticle, from [186].**

Central nervous system (CNS) consists of the brain and the ventral nerve cord.

Metabolites enriched in the CNS are part of amino acid metabolism, thiamine metabolism and insect hormone biosynthesis (Table 5-4). The nervous system is the major site for the synthesis and signalling molecules including hormones and neuropeptides. Hence, it makes sense that the tissue is enriched in amino acids and insect hormones.

Unfortunately, it is not possible to compare the two datasets. Even though, the brain is present in both samples, comparison of the whole CNS with the head would not be a viable one. As is clear from the data obtained from SU, majority of enriched metabolites in the head represent visual pigments, which might ‘mask’ compounds related to the brain and the nervous system. This points out how important it is to study individual tissues, and group samples according to their function, i.e. CNS tissues together, rather than the whole head.



**Table 5-4 Metabolites enriched in central nervous system (the highest tissue enrichment value in CNS). Results are represented as a ratio of relative abundance in a given tissue in comparison to other samples. Metabolite levels are coloured according to relative intensity (blue = low, red = high, blue<yellow<orange<red).**

Mass	RT	Formula	Putative metabolite	Pathway	CNS	Gut	MTs	WF
143.04	8.33	C <sub>6</sub> H <sub>9</sub> NOS	5-(2-Hydroxyethyl)-4-methylthiazole	Thiamine metabolism	93.05	0.00	0.00	0.00
210.08	5.82	C <sub>13</sub> H <sub>10</sub> N <sub>2</sub> O	2-Aminoacridone	No known pathway	61.63	0.00	0.00	0.00
167.06	5.72	C <sub>8</sub> H <sub>9</sub> NO <sub>3</sub>	3-Methoxyanthranilate	Tryptophan metabolism	36.98	0.00	0.00	0.00
175.06	6.02	C <sub>10</sub> H <sub>9</sub> NO <sub>2</sub>	Indole-3-acetate	Tryptophan metabolism	35.87	7.25	15.23	0.00
251.10	12.96	C <sub>10</sub> H <sub>13</sub> N <sub>5</sub> O <sub>3</sub>	Deoxyadenosine	Purine metabolism	5.21	1.18	0.94	1.00
297.09	11.04	C <sub>11</sub> H <sub>15</sub> N <sub>5</sub> O <sub>3</sub> S	5'-Methylthioadenosine	Arginine and proline metabolism	7.10	1.60	2.19	1.00
158.06	5.71	C <sub>7</sub> H <sub>10</sub> O <sub>4</sub>	2-Isopropylmaleate	Valine, leucine and isoleucine biosynthesis	2.77	0.38	0.97	1.00
148.05	5.79	C <sub>9</sub> H <sub>8</sub> O <sub>2</sub>	trans-Cinnamate	Phenylalanine metabolism	21.67	16.07	12.99	0.00
190.08	5.81	C <sub>8</sub> H <sub>14</sub> O <sub>5</sub>	(R)-3-((R)-3-Hydroxybutanoyloxy)butanoate	Butanoate metabolism	8.71	5.32	5.45	1.00
446.34	5.26	C <sub>28</sub> H <sub>46</sub> O <sub>4</sub>	3-Dehydroteasterone	Brassinosteroid biosynthesis	3.95	3.07	1.65	1.00
270.18	5.64	C <sub>15</sub> H <sub>26</sub> O <sub>4</sub>	(10S)-Juvenile hormone III acid diol	Insect hormone biosynthesis	2.61	1.73	1.54	1.00
282.18	5.64	C <sub>16</sub> H <sub>26</sub> O <sub>4</sub>	12-trans-Hydroxy juvenile hormone III	Insect hormone biosynthesis	2.70	1.86	1.83	1.00
141.04	6.20	C <sub>6</sub> H <sub>7</sub> NO <sub>3</sub>	6-oxo-1,4,5,6-tetrahydronicotinate	Nicotinate and nicotinamide metabolism	7.10	9.04	4.17	1.00

Altogether these results confirmed that tissue metabolomes differ from the whole fly and from each other. These differences seem to reflect functional differences between individual tissues and provide new hypotheses about tissue-tissue interactions. However, some metabolites and metabolic pathways are conserved among tissues. Moreover, the results validated the method for the study of individual *Drosophila* tissues. Based on these findings, we studied purine metabolism mutants using whole flies as well as individual tissues.

## **5.2.2 Metabolomics analysis of mutants of purine metabolism**

Two different sets of experiments were performed: one using adult flies and another one using L3 larva. Experiments in the adult consisted of whole flies and Malpighian tubules. The fly lines used were BDSC 18554, and BDSC 17767 flies. After performing these experiments, we realised that it would be more informative to use larval tissues and include BDSC 18814 (*UO* mutant) flies. Hence, larval experiments were performed using L3 Malpighian tubules and whole larvae, and the fly lines used were BDSC 18554, BDSC 18814 and BDSC 17767 flies. Complete data sets can be viewed in Appendix VII and VIII.

### **5.2.2.1 Adult fly metabolomes**

210 metabolites were identified in whole fly samples, and 146 metabolites in adult Malpighian tubules. The instrument settings as well as data analysis are described in Materials and Methods (Chapter 2.16). The relative abundance of the experimental group (BDSC18554) was expressed as explained in chapter 5.2.1 relative to the mean signal intensity of the control group (BDSC17767). Metabolite levels were coloured according to relative abundance (blue = low, red = high) (Appendix VII and VIII). Moreover, t-test was performed and where the P-value (from t-test against control group) was less than 0.05, the data was analysed further. Based on our hypothesis that *CG30016* is a 5-hydroxyisourate hydrolase, we expected to see differences in purine metabolism between the mutant and the control. If *CG30016* were a 5-HIU hydrolase, metabolites upstream from the knockout would be enriched and the ones downstream would be decreased.

In both whole fly and MTs, purine metabolites were detected and were among the most abundant metabolites in the mutant (Tables 5-5 and 5-6). Metabolites involved in numerous other pathways were also detected and enriched in both whole fly and MTs. It

is interesting to see that a single gene knock out in the fly affects several unrelated pathways (Figure 5-8). Among detected metabolites, the majority were part of lipid and amino acid metabolism. However, this does not reflect the differences between the mutant and the control. This could be because the detection method is more sensitive to these metabolites or simply because there are more amino acids and lipids in the samples than other type of compounds.

**Table 5-5 The most abundant metabolites in whole *CG30016* mutant flies. Metabolite names in bold are identifications, for which standards were available and run alongside experimental samples.**

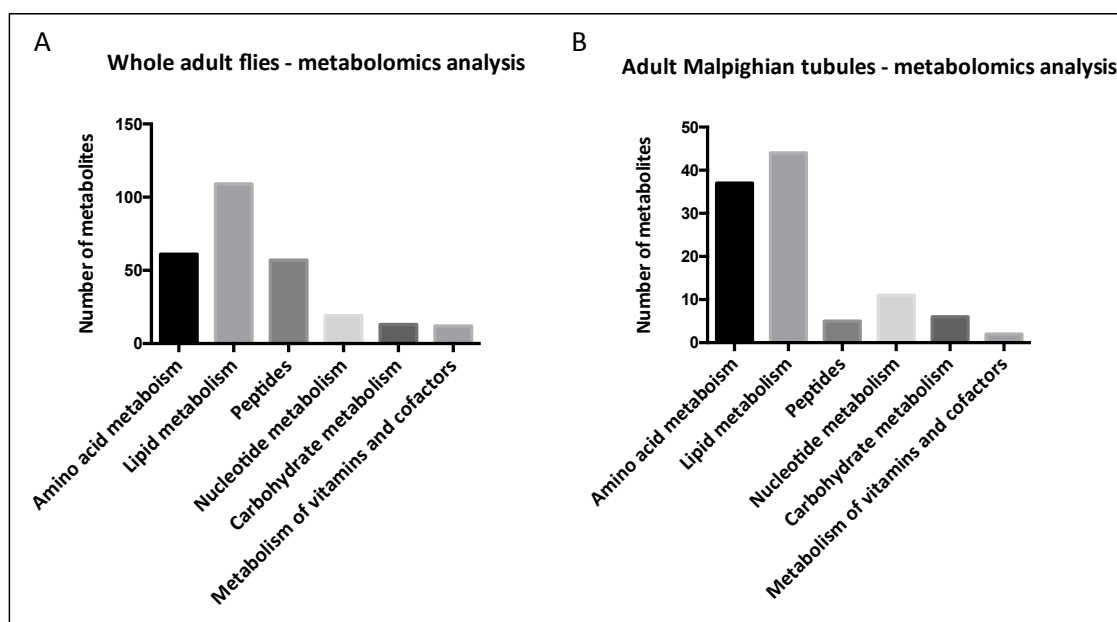
Metabolite name	Map	KEGG Pathway	P-value	BDSC 18554 (mutant)
Xanthurenic acid	Amino acid metabolism	Tryptophan metabolism	0.001	28.31
2-Methoxyhexadecanoic acid	Lipid metabolism	Fatty acids and conjugates	0.031	6.82
D-Sorbitol	Carbohydrate metabolism	Fructose and mannose metabolism	0.023	6.38
<b>Imidazole-4-acetate</b>	Amino acid metabolism	Histidine metabolism	0.040	4.08
Guanosine	Nucleotide metabolism	Purine metabolism	0.008	3.31
<b>L-Citrulline</b>	Amino acid metabolism	Arginine and proline metabolism	0.001	2.96
<b>Inosine</b>	Nucleotide metabolism	Purine metabolism	0.009	2.58
Imidazole-4-acetaldehyde	Amino acid metabolism	Histidine metabolism	0.002	2.52
<b>4-Trimethylammoniobutanoate</b>	Amino acid metabolism	Lysine degradation	0.013	2.28
<b>Xanthine</b>	Nucleotide metabolism	Purine metabolism	0.049	2.01

**Table 5-6 The most abundant metabolites in Malpighian tubules of adult *CG30016* mutant flies. Metabolite names in bold are identifications, for which standards were available and run alongside experimental samples.**

Metabolite name	Map	KEGG Pathway	P-value	BDSC 18554 (mutant)
5-Hydroxy-2-oxo-4-ureido-2,5-dihydro-1H-imidazole-5-carboxylate (OHCU)	Nucleotide metabolism	Purine metabolism	0.036	5.751
O-Propanoylcarnitine	Lipid metabolism	Oxidation of branched fatty acids	0.044	4.51
<b>D-Glucose 6-phosphate</b>	Carbohydrate metabolism	Starch and sucrose metabolism	0.011	3.38
<b>L-Carnitine</b>	Amino acid metabolism	Lysine degradation	0.010	2.98
Glutathione disulfide	Amino acid metabolism	Glutathione metabolism	0.018	2.31
Urate	Nucleotide metabolism	Purine metabolism	0.049	2.25
Allantoin	Nucleotide metabolism	Purine metabolism	0.042	1.83
5-Hydroxyisourate	Nucleotide metabolism	Purine metabolism	0.046	1.52
Guanosine	Nucleotide metabolism	Purine metabolism	0.034	1.48
L-Kynurenine	Amino acid metabolism	Tryptophan metabolism	0.012	1.44

The top 10 metabolites enriched in whole knockout flies are illustrated in Table 5-5. Among the most abundant metabolites, three are part of purine metabolism: guanosine, inosine and xanthine. Interestingly, the single most abundant metabolite is xanthurenic acid, which is a part of tryptophan metabolism. On the other hand, in the tubules, the single most abundant metabolite is OHCU (Table 5-6). Moreover, other metabolites involved in urate degradation are enriched in the mutant tubule, including urate, 5-HIU and allantoin.

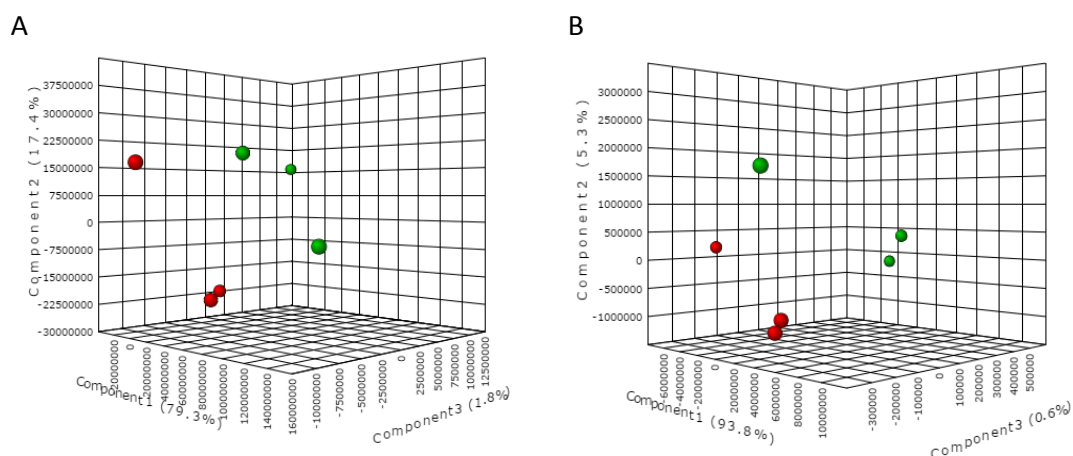
There are clear differences between the whole fly and the MTs analyses. Some purine metabolism metabolites, including 5-HIU and OHCU, were not detected in the whole fly at all. Moreover, the results for certain metabolites are very different in the whole fly and MTs. This could be due to metabolites present in other tissues ‘masking’ the results specific to MTs and purine metabolism. Since the urate degradation pathway takes place in the MTs and *CG30016* is expressed almost exclusively in the tubule, it makes more sense to use individual tissue for this kind of analysis rather than the whole fly. This way the method is more sensitive and provides more details on the pathway of interest.



**Fig. 5-8 Metabolic pathways affected by the knockout of *CG30016* in whole flies.**

PCA analysis was performed using MetaboAnalyst (<http://www.metaboanalyst.ca>), an open-source platform for metabolomics data analysis [258, 458]. The analysis was carried out in order to determine how varied the replicates were within each sample, whether they clustered together, and whether there were differences between the control and

the mutant samples. The results are illustrated in Figure 5-9. In both the whole fly (Figure 5-9A) and the MTs (Figure 5-9B), control samples (red) clustered away from mutant samples (green). Moreover, there was variability between replicates within both samples.

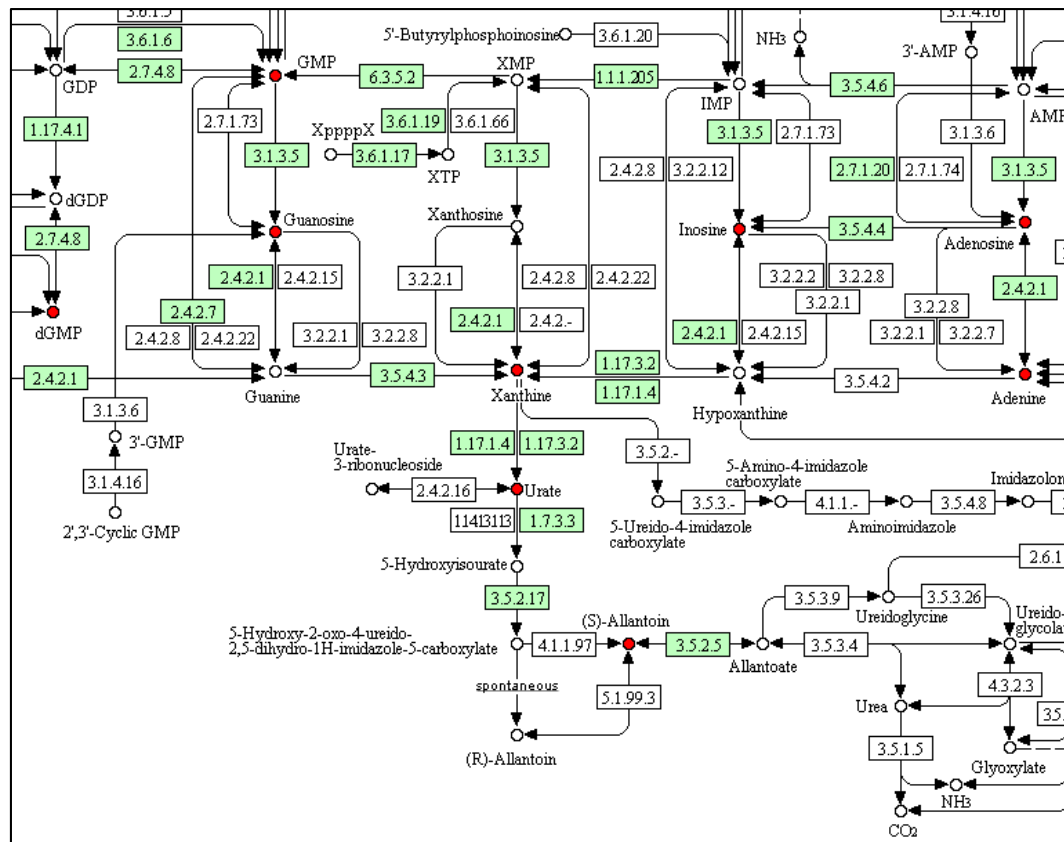


**Fig. 5-9 Results of the PCA analysis: red dots represent control sample replicates (BDSC 17767), and green dots represent mutant sample replicates (BDSC 18554). (A) PCA of the whole fly samples, (B) and Malpighian tubule samples. In Both the whole fly and MTs, replicates of each sample are more related to each other than the other sample. Generated using MetaboAnalyst (<http://www.metaboanalyst.ca>).**

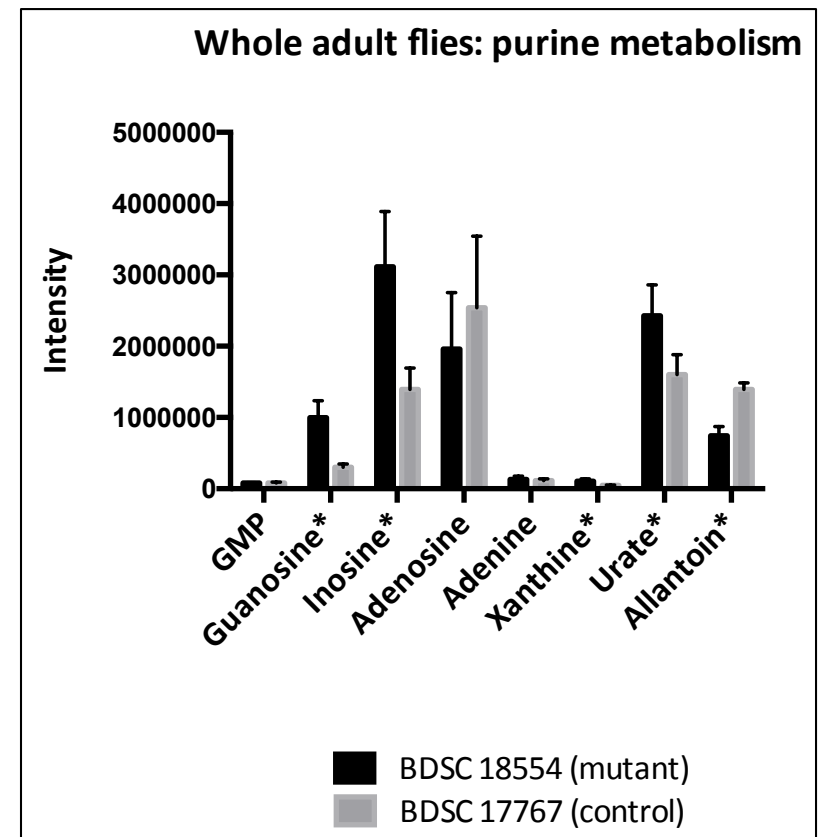
#### 5.2.2.2 Adult fly metabolomes – purine metabolism

Within purine metabolism 9 (Malpighian tubules) and 8 (whole fly) metabolites with different abundance in the control and the mutant were detected (Figures 5-10A and 5-11A). Whereas in the whole fly most abundant purine metabolites were not directly involved in 5-HIU hydrolase reaction, in MTs the four most abundant metabolites were urate, 5-HIU, OHCU and allantoin. Moreover, the single most abundant metabolite in the tubule was OHCU (Table 5-6), which is a product of the reaction catalysed by 5-HIUH. OHCU concentration in *CG30016* mutant tubules was increased 5.75-fold when compared to the control. This is opposite to what we expected. It is also surprising to see that the abundance of allantoin is higher in the mutant than the control MTs. On the other hand, as predicted, 5-HIU and other metabolites upstream from 5-HIU hydrolase were enriched in the mutant (Figure 5-10B and 5-11B). The only exception is adenine, which is enriched in the control.

A



B



**Fig. 5-10 Changes in purine metabolism in the whole fly caused by the knockout of *CG30016*.** (A) KEGG map of *Drosophila melanogaster* purine metabolism; red circles represent metabolites that were identified by the LC-MS analysis. (B) Abundance of metabolites involved in purine metabolism compared between the knockout and the control flies. Asterisk indicates statistically significant differences. The abundance is represented as the signal intensity.



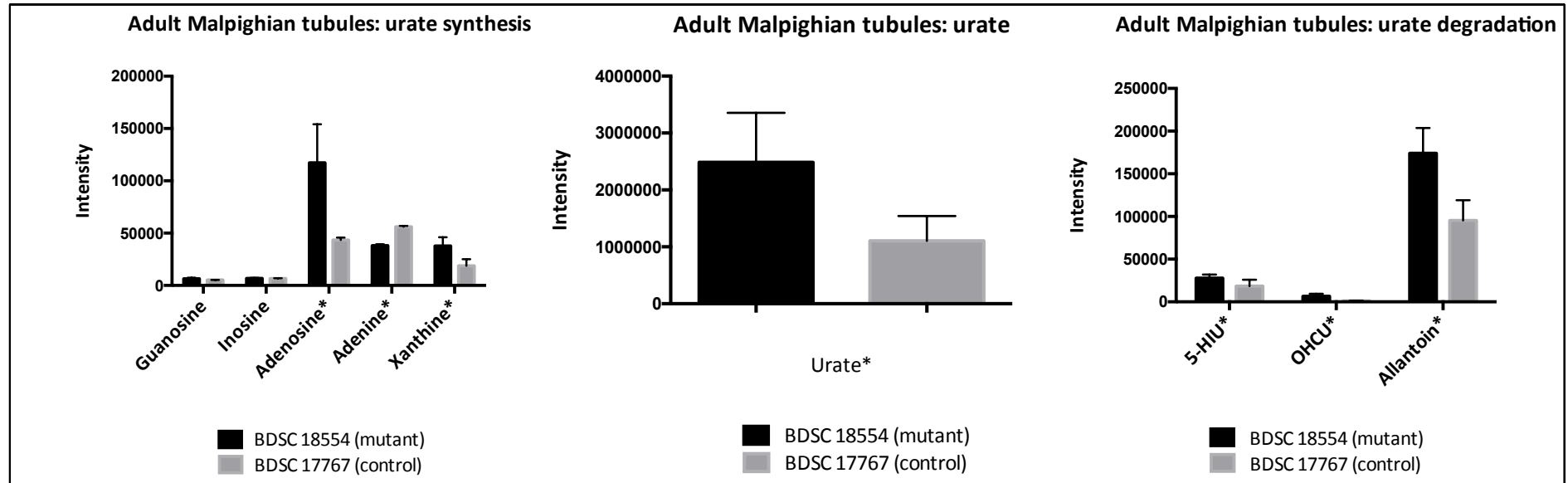
The results in the whole fly differ from the ones in the MTs. For example, the abundance of allantoin is higher in the mutant tubule than the control (Figure 5-11B). In the whole fly on the other hand, allantoin is enriched in the control when compared to the mutant (Figure 5-10B). This might reflect the differential distribution of metabolites across the tissues. Perhaps the concentration of allantoin increased in mutant tubules but not other tissues resulting in lower levels of allantoin in the whole fly. Once again, this confirms that it is more informative to perform metabolomics analysis in the tissue of interest rather than the whole organism.

Interestingly, allantoin has two isoforms and it is impossible to distinguish between them using LC-MS. In fact one of the isoforms (R)-allantoin is a product of non-enzymatic degradation of OHCu, and (S)-allantoin is a product of OHCu conversion catalysed by OHCu decarboxylase [355, 461]. Here, it is impossible to tell, which isoform is enriched in the tubules and the whole fly.

The diagram illustrates the metabolic pathways of purine metabolism. It shows the conversion of XppppX to AMP and the degradation of AMP to CO2 and NH3. Key intermediates include Guanine, Xanthosine, Xanthine, Inosine, Adenosine, Adenine, Hypoxanthine, Urate, 5-Hydroxyisourate, (S)-Allantoin, Allantoate, Ureidoglycine, Ureidoglycolate, Urea, Glyoxylate, and NH3. The final products are CO2 and NH3. The map is color-coded: green boxes represent genes, white boxes represent metabolites, and red circles represent reactions. The map is organized into three main sections: the top section shows the conversion of XppppX to AMP, the middle section shows the degradation of AMP to Urate, and the bottom section shows the degradation of Urate to CO2 and NH3.

**Fig. 5-11 Changes in purine metabolism in adult Malpighian tubules caused by the knockout of *CG30016*. (A) KEGG map of *Drosophila melanogaster* purine metabolism; red circles represent metabolites that were identified in adult Malpighian tubules by the LC-MS analysis.**

B



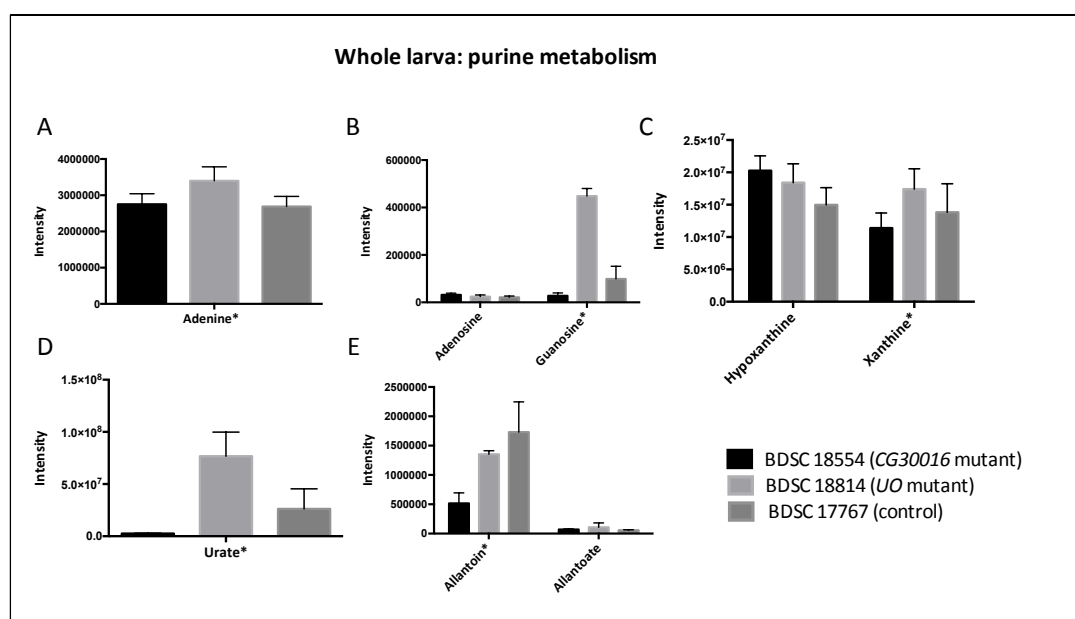
**Fig. 5-11 Changes in purine metabolism in adult Malpighian tubules caused by the knockout of *CG30016*. (B) Abundance of metabolites involved in purine metabolism compared between the knockout and the control flies; (left) metabolites involved in urate synthesis (upstream from 5-HIU hydrolase and predicted function of *CG30016*), (middle) abundance of urate in both fly lines, (right) abundance of metabolites downstream from urate, involved in urate degradation. Asterisk indicates statistically significant differences. The abundance is represented as the signal intensity.**

Our results clearly indicate that *CG30016* is involved in purine metabolism and specifically urate degradation pathway. However, it is impossible to conclude whether *CG30016* is a 5-HIU hydrolase. In fact there are several possible conclusions based on the results. One of the possibilities is that the enrichment of OHCU in the mutant tubule reflects decreased activity of OHCU decarboxylase. This would indicate that the *CG30016* mutant is the OHCU decarboxylase rather than a 5-HIU hydrolase knockout, and that *CG30016* codes for OHCU decarboxylase. This is unlikely because *CG30016* has no sequence similarity to other OHCU decarboxylases and has no decarboxylase domain.

An alternative explanation is that *CG30016* is a 5-HIU hydrolase and its knockout can be observed as 5-HIU enrichment in the mutant tubules. However, since 5-HIU can degrade spontaneously to OHCU and allantoin, both metabolites are also enriched in the tubule. However, because LC-MS cannot differentiate between different isoforms of allantoin, it is impossible to fully support this hypothesis.

#### **5.2.2.3 Mutant larva metabolomes**

Metabolomics analysis was performed using whole L3 larvae as well as larval MTs, and was carried out by myself at SU facility. Metabolites were identified and quantified using Xcalibur software combined with Sieve Software 1.3 (Thermo Fisher Co.). The abundance of metabolites was calculated and expressed as intensity, and compared between the *CG30016* and *UO* mutant and the control. Except metabolites that matched masses and retention times of standards, all metabolites were putatively identified. All masses were within 1.5 ppm of the exact molecular formula. This means that the only competing metabolites within the database search were isomers. A t-test was performed to establish statistical significance, and where the P-value was below 0.05, the data were analysed further. In both cases there were differences in metabolite abundance within purine metabolism between mutants and the control (Figure 5-12) (for original data see Appendices X, XI and XII). Unfortunately, the metabolites directly upstream and downstream from 5-HIU hydrolase (5-HIU and OHCU) were not detected. This could be because they are both unstable as opposed to urate and allantoin that were detected.



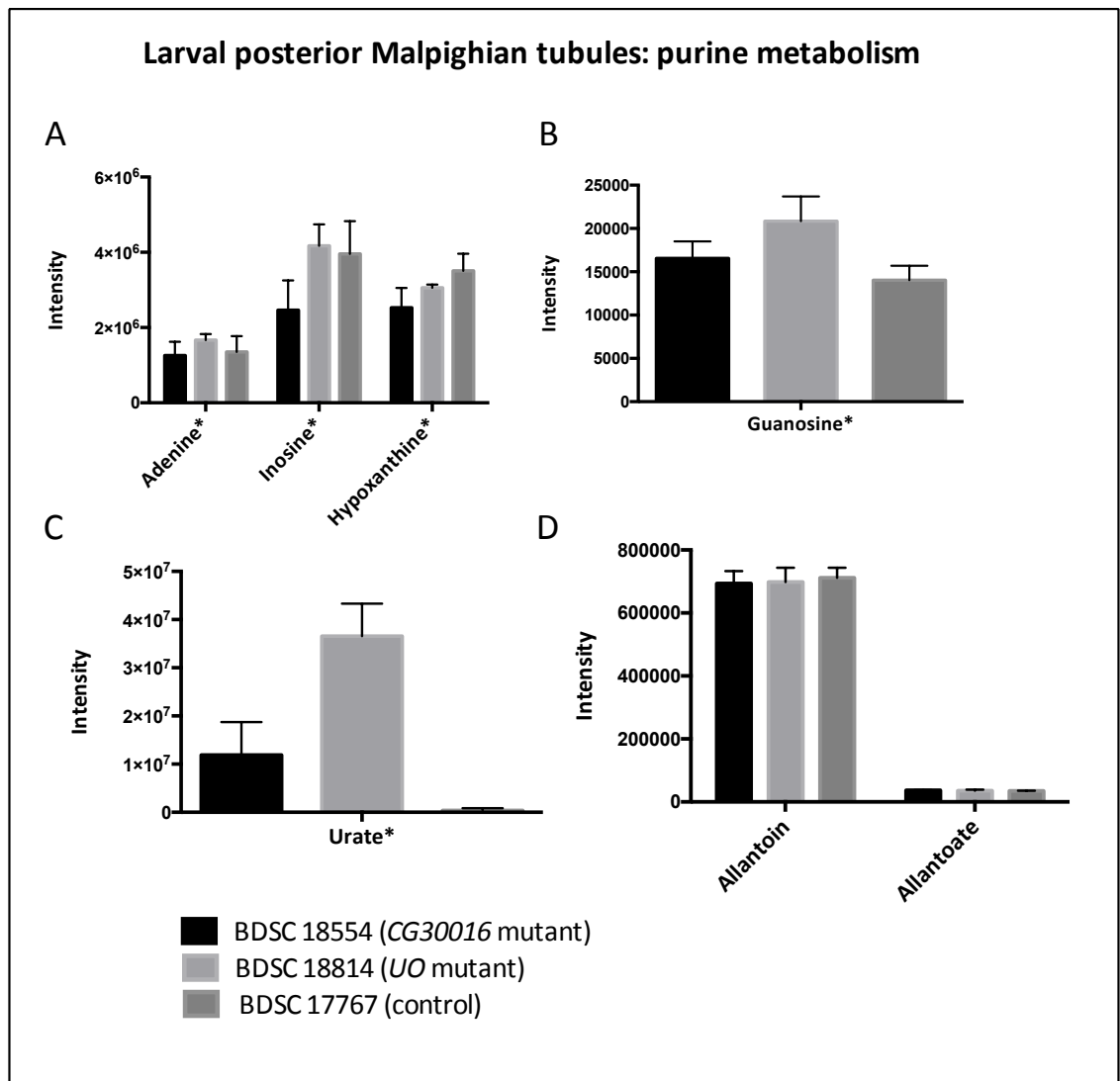
**Fig. 5-13** Changes in purine metabolism in whole larvae caused by the knockout of *UO* and *CG30016*. Abundance of metabolites involved in purine metabolism compared between the knockouts and the control flies is represented as the signal intensity. Asterisk indicates statistically significant differences. (A-E) Metabolites are represented in several graphs depending on the signal intensity. Metabolites upstream from *UO* and predicted *CG30016* are illustrated in figures A-D, and the ones downstream from the lesion are illustrated in figure E.

The analysis was also performed for larval tubules. It was carried out separately for anterior and posterior tubules because of an observation in Chapter 3.2.1 that there are more urate crystals in the posterior tubule of L3 larvae.

In the *UO* mutant, levels of most metabolites upstream from *UO* were higher in the mutant posterior tubule than in the control (Figure 5-14A-C). The only exception was hypoxanthine, which is higher in the control. The greatest difference is urate, which is highly enriched in the mutant (Figure 5-14C). Similar results were observed in the anterior tubule. However, the differences between the mutant and the control were smaller than in the posterior tubule. This is consistent with the results in Chapter 3.2.1, where less urate crystals were observed in the anterior tubule. Moreover, there were no significant differences in the abundance of adenosine, guanosine, inosine and hypoxanthine between the mutant and the control. Altogether, these results confirm the role of *UO* in *Drosophila* tubules.

In both anterior and posterior tubules there were no significant differences in the abundance of allantoin and allantoate between the mutant and the control. This is surprising as we expected to observe a pattern similar to the whole larvae, with

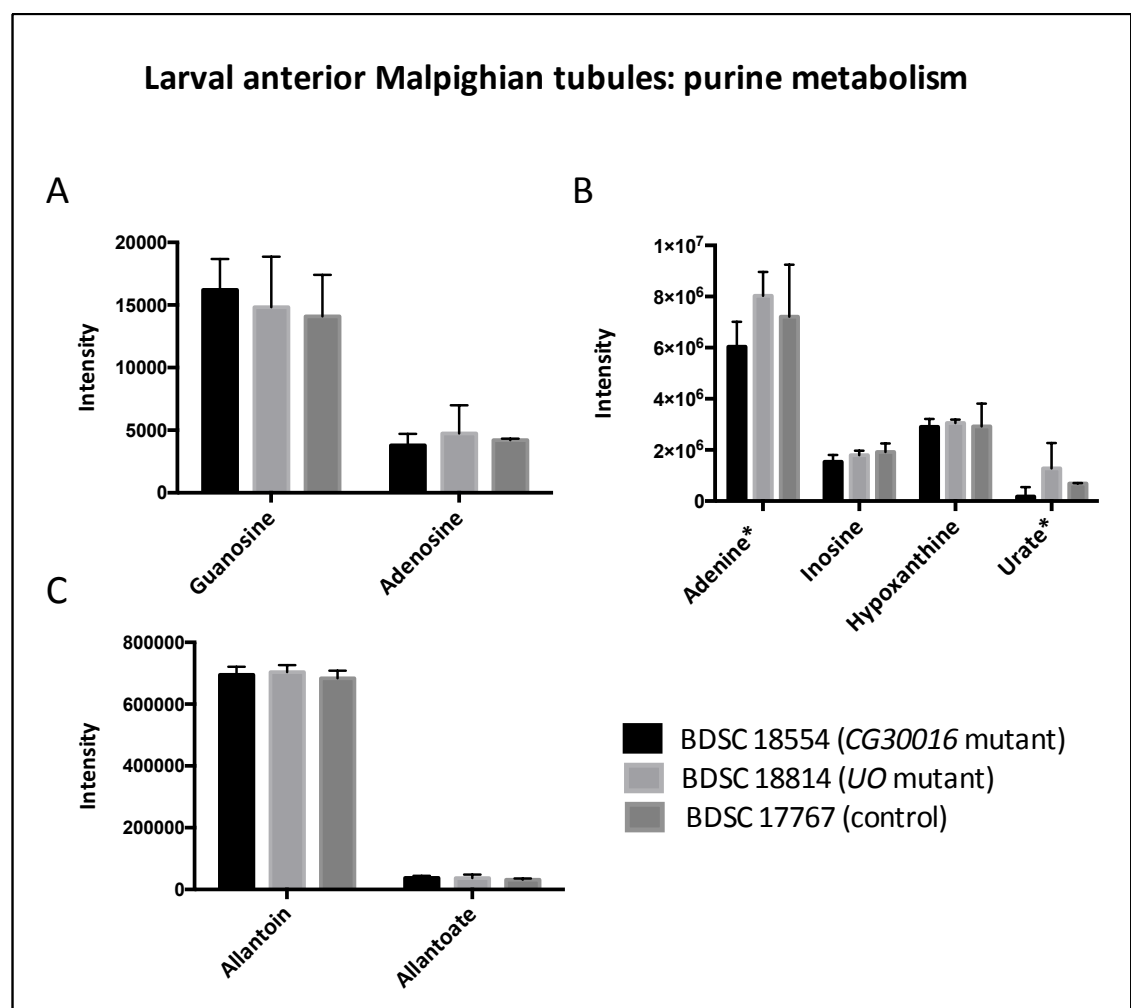
decreased allantoin in the mutant.



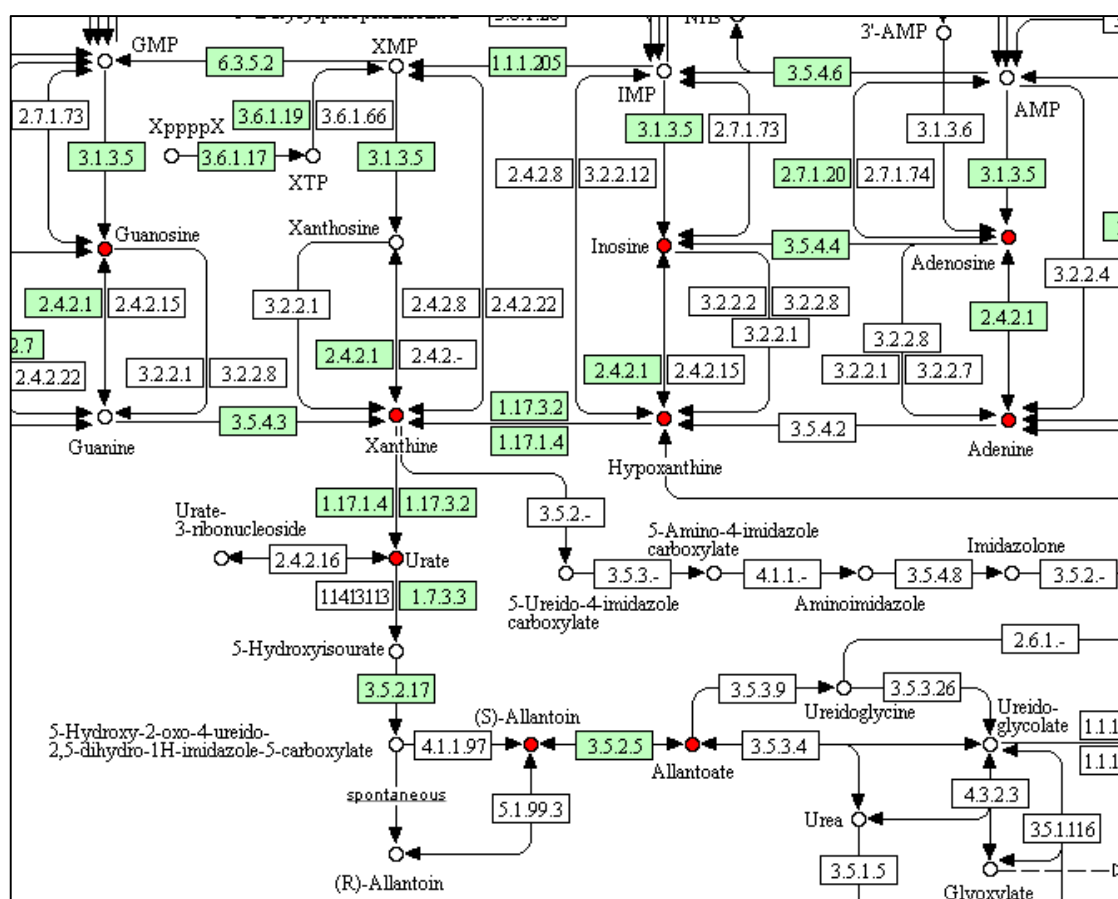
**Fig. 5-14** Changes in purine metabolism in larval posterior Malpighian tubules caused by the knockout of *UO* and *CG30016*. Abundance of metabolites involved in purine metabolism compared between the knockouts and the control flies is represented as the signal intensity. Asterisk indicates statistically significant differences. (A-D) Metabolites are represented in several graphs depending on the signal intensity. Metabolites upstream from *UO* and predicted *CG30016* are illustrated in figures A-C, and the ones downstream from lesions are illustrated in figure D.

In terms of the *CG30016* mutant, the results are also different between the posterior and anterior tubule. In the posterior tubule, urate is more abundant in the mutant than in the control, and so is guanosine. However, the levels of adenine, inosine and hypoxanthine are lower in the knockout than in the control (Figure 5-15A-C). This is surprising and might not be a direct result of *CG30016* knockout but rather a difference between individual flies. In the anterior tubule, both urate and adenine are less abundant in the mutant than the control and there are no statistically significant differences for other metabolites. In

both posterior and anterior tubules there are no difference in the level of allantoin and allantoate between the mutant and the control. These results suggest that *CG30016* is involved in purine metabolism and urate degradation pathway. However, it might play a different role or be expressed at different levels in anterior and posterior tubules. One of the reasons for this could be a functional difference between the anterior and posterior tubule. For example, the anterior tubule might not contribute to urate degradation pathway as much as the posterior tubule. Hence, more urate crystals were observed in the posterior tubule (Chapter 3.2.1), as well as greater effects of *UO* and *CG30016* knockout on the levels of urate degradation metabolites in the posterior tubule.



**Fig. 5-15** Changes in purine metabolism in larval anterior Malpighian tubules caused by the knockout of *UO* and *CG30016*. Abundance of metabolites involved in purine metabolism compared between the knockouts and the control flies is represented as the signal intensity. Asterisk indicates statistically significant differences. (A-C) Metabolites are represented in several graphs depending on the signal intensity. Metabolites upstream from *UO* and predicted *CG30016* are illustrated in figures A-B, and the ones downstream from lesions are illustrated in figure C.



In whole larva, metabolites upstream from UO, were increased in BDSC 18814 mutant (Figure 5-13A-D). The biggest difference between the mutant and the control was urate, which was highly enriched in BDSC 18814 flies. This confirms that the mutant is a *UO* knockout. Moreover, the abundance of allantoin was lower in the mutant than in the control, which is consistent with previous results. There was no significant difference between the levels of allantoate in the mutant and the control (Figure 5-13E).



Altogether, these results confirm the role of *Drosophila UO* gene and suggest that *CG30016* is involved in urate degradation pathway. However, the role of *CG30016* as 5-HIU hydrolase has not been confirmed. Unfortunately, 5-HIU and OHCU were not detected in the analysis. Once again, this could be because both metabolites are unstable and might not always be detected in LC-MS.

### **5.2.3 Metabolomes of secreted fluid**

In this experiment 364 metabolites were identified using methods described in Chapter 2.16. Metabolite levels for each experimental group were expressed as mean peak intensity (height) relative to the mean peak intensity of the control group (Schneider's medium). Metabolite levels are coloured according to relative intensity (blue = low, red = high) (Appendix IX). Moreover, t-test was performed and where the P-value (from t-test against control group) is less than 0.05, the intensity is highlighted in bold in the original spreadsheet.

Metabolites identified in the secreted fluid are part of several pathways including purine and tryptophan metabolism (Table 5-7). This is consistent with our previous results showing that the tubule is enriched in tryptophan-related metabolites as well as members of the purine metabolism.

**Table 5-7 The most abundant metabolites in secreted fluid. Metabolite levels are expressed as mean peak intensity relative to the mean peak intensity of the control group (Schneider's medium). Metabolite names highlighted in bold are chemicals matched to standards by exact mass and retention time.**

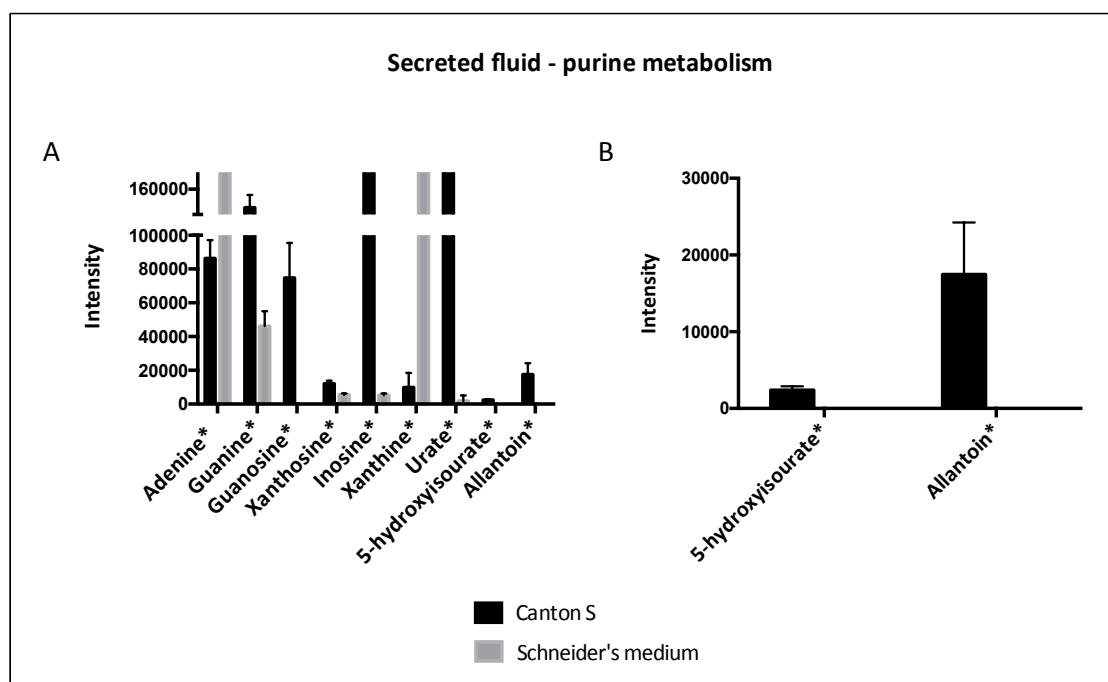
Metabolite name	Map	KEGG Pathway	Schneider's medium (control)	Canton S (sample)
<b>L-Kynurenine</b>	Amino acid metabolism	Tryptophan metabolism	0.00	697.73
3-Hydroxy-L-kynurenine	Amino acid metabolism	Tryptophan metabolism	0.00	196.37
Urate	Nucleotide metabolism	Purine metabolism	1.00	156.89
Xanthurenic acid	Amino acid metabolism	Tryptophan metabolism	0.00	125.81
<b>Guanosine</b>	Nucleotide metabolism	Purine metabolism	0.00	74.66
<b>D-Glucose 6-phosphate</b>	Carbohydrate metabolism	Starch and sucrose metabolism	0.00	71.29
5-Hydroxyindoleacetate	Amino acid metabolism	Tryptophan metabolism	0.00	63.71
S-Adenosyl-L-methionine	Amino acid metabolism	Arginine and proline metabolism	0.00	56.43
<b>Inosine</b>	Nucleotide metabolism	Purine metabolism	1.00	49.25
3'-AMP	Nucleotide metabolism	Purine metabolism	1.00	31.57
4-Acetamidobutanoate	Amino acid metabolism	Arginine and proline metabolism	1.00	26.63

Cys-Thr-Ser-Ser	Peptides	Polar peptide	0.00	23.01
2-Methyl-1-hydroxypropyl-ThPP	Amino acid metabolism	Valine, leucine and isoleucine degradation	0.00	21.72
2-Aminophenol	Amino acid metabolism	Tryptophan metabolism	0.00	20.77
Phe-Ala-Gly	Peptides	Hydrophobic peptide	0.00	18.20
(S)(+)-Allantoin	Nucleotide metabolism	Purine metabolism	0.00	17.45
L-Formylkynurenine	Amino acid metabolism	Tryptophan metabolism	1.00	14.79
<b>Choline phosphate</b>	Lipid metabolism	Glycerophospholipid metabolism	1.00	13.97
<b>L-2-Aminoadipate</b>	Amino acid metabolism	Lysine biosynthesis	0.00	10.85
Methylglyoxal	Amino acid metabolism	Glycine, serine and threonine metabolism	0.00	10.24

The experiment was performed using secreted fluid of wild-type Canton S flies. The main focus was on the purine metabolism in order to test the method as a suitable one to study purine metabolism mutants. Hence, we looked at purine metabolism metabolites to test whether they are detected in the secreted fluid. Figure 5-16 shows that numerous metabolites involved in purine metabolism were identified. Among these, most were significantly enriched in the secreted fluid in comparison to the control. The exceptions were adenine and xanthine, which were enriched in the Schneider's medium (Figure 5-16A).  $\beta$ -Alanine is one of the ingredients found in Schneider's medium and so it is likely that the results reflect this. However, it is not clear why xanthine is enriched in the

medium but not the secreted fluid.

Altogether these results showed that LC-MS is a suitable technique not only for tissues and whole organism studies but also secreted fluid. It also validated the method for the study of *Drosophila melanogaster* mutants described in previous Chapter.



**Fig. 5-16 Metabolites involved in purine metabolism enriched in secreted fluid in comparison to Schneider's medium. Abundance of metabolites is represented as the signal intensity. Asterisk indicates statistically significant differences. (A) All metabolites involved in purine metabolism are illustrated. (B) Metabolites involved in urate degradation pathway and downstream from urate are illustrated.**

### 5.3 Discussion

Our results show that tissue metabolomes differ significantly from each other and from the whole fly. These differences seem to be correlated to tissue function and require further analysis. In the future it would be interesting to perform tissue metabolomics using more than four replicates, since we observed a lot of variability between replicates. These results are consistent with FlyAtlas findings that tissue transcriptomes differ and the differences are functionally significant [4]. However, it is important to note that some metabolites and metabolic pathways are conserved among tissues, which might reflect the fact that some organismal functions are not tissue-specific. Moreover, we validated LC-MS based metabolomics as a suitable method to study individual tissues and changes between flies with different genotypes. It is also clear that comparing two different data

sets performed at different times and different facilities requires a standard protocol. Most importantly, it has been noted here that the comparison is possible only when the samples contain the same tissues (i.e. the head sample cannot be compared to CNS, mid gut cannot be compared to the whole gut etc.).

Our results confirmed that knocking out *UO* in the fruit fly results in build up of urate and upstream metabolites and decrease of downstream metabolites. Moreover, more urate was detected in the posterior than anterior tubule. This is consistent with our previous observation that more urate crystals are visible in the posterior tubule of the mutant. This validates the method and confirms it is suitable for the detection of changes in metabolite abundance as a result of a single-gene mutation. Moreover, it revealed that single tissue metabolomics is more informative in this case than whole fly analysis.

Following on from this finding, metabolomics analysis of urate degradation mutants was performed. These results confirm that *CG30016* is involved in urate degradation pathway. It is clear that, in MTs of *CG30016* mutant urate degradation is blocked because of build up of metabolites upstream from expected knockout and decrease of downstream metabolites. However, it is impossible to establish its exact role in the pathway based on the results. Moreover, our results of mutant tubules revealed that single-gene mutations not only influence levels of metabolites next to a knockout, but also affect numerous unrelated metabolic pathways. This highlights how small changes affect hundreds of reactions in an organism. In order to determine the role of *CG30016* in the fruit fly, further research is required. Following hypotheses would be tested:

1. *CG30016* is a 5-HIU hydrolase and the increased 5-HIU in knockout tubules result in the increase of OHCU and allantoin, which is the outcome of the non-enzymatic reaction. In order to test this hypothesis, the differentiation between different isoforms would be required, which is currently not possible with available equipment.
2. *CG30016* mutant is OHCU decarboxylase. This is unlikely because *CG30016* has no sequence similarity to other OHCU decarboxylases and has no decarboxylase domain. However, the gene could code for a novel decarboxylase, which is not similar to other OHCU decarboxylases but catalyses the same reaction. Once again, differentiation between allantoin isoforms would help answer this question.
3. *Drosophila CG30016* is a bifunctional enzyme playing a role of both 5-HIUH and OHCU decarboxylase. This phenomenon was observed in some organisms (*A. thaliana* and

*Bacillus subtilis*), in which two enzymes are fused into a single polypeptide [102, 462-464]. However, the fact that *CG30016* is missing a decarboxylase domain makes this hypothesis unlikely.

4. There are two 5-HIU hydrolases and only one of them is encoded by *CG30016*. It forms a complex with OHCU decarboxylase and is essential for its function so when it is missing, OHCU is not converted to allantoin and builds up. The other enzyme provides enough activity to convert 5-HIU to OHCU and is not encoded by *CG30016*, which results in OHCU built up.
5. There is an alternative pathway for the breakdown of urate to allantoin.

Moreover, the technique has several limitations, which we observed during these experiments. Solving them would increase the chance to identify the role of *CG30016*. For example, it would be useful to obtain all the standards for urate degradation intermediates. This is particularly challenging because 5-HIU and OHCU are unstable. Another problem we encountered was no detection of 5-HIU and OHCU in larval samples. Once again, this is probably due to their instability. It is necessary in this case to come up with a solution or a technique that would overcome this problem. Perhaps, using another column or combination of different columns for the LC would improve the detection of these metabolites. Finally, developing a technique to distinguish between allantoin isoforms would allow distinguishing between the spontaneous and enzymatic pathways of urate degradation. For example CD-spectroscopy can be used to obtain spectra of allantoin and determine its configuration.

The analysis of secreted fluid showed that LC-MS is not only suitable for the analysis of individual fly tissues but also secreted fluid. Moreover, metabolites involved in purine metabolism were detected, including 5-HIU and OHCU. It would be interesting to perform this experiment using secreted fluid from urate degradation mutant tubules. It might be easier to detect the unstable metabolites than in the whole tissue.

Finally, general limitations of metabolomics presented in Chapter 1.3.5 are the main challenge in our work, including metabolite identification and quantification, data interpretation, data visualisation and integration of metabolomics data, and data obtained from other omics. Our results showed that metabolomics is very complex and requires time-consuming analysis. It is a fantastic method to generate new hypotheses and provide a big picture of metabolic state of a given sample. However, it is quite

challenging when looking at individual pathways and metabolites and in order to provide answers to hypotheses. Statistical analysis of metabolomics data presented a great challenge, especially in case of tissue-specific metabolomes, where multiple study groups were compared. It stresses the importance of biostatistics and bioinformatics expertise in case of metabolomics studies. Selecting the correct statistical method and visualisation tools is crucial for data interpretation of LC-MS results. Collaboration with a biostatistician would be beneficial here in order to perform more statistical analyses and provide better data presentation. A great example of untargeted metabolomics data analysis is presented in the following study. Wikoff *et al.* used unequal-variance t-tests and fold-change to identify statistically significant differences between cerebrospinal fluid of SIV-infected and uninfected macaques [465]. As a result out of 3000 measured features, 12 were selected and further identified and analysed. This study shows a different analysis approach, where statistical analysis precedes the identification step. This reduces the need for painstaking data interpretation of huge datasets. However, in case of our results, we aimed to compare as many features as possible (in case of tissue-specific study), or looked at specific pathways (in urate degradation knockout study) so a different type of analysis is required.

Moreover, constant developments in metabolomics technologies, as well as data processing tools and algorithms, require regular updating of the current tissue-specific metabolome of the fruit fly. Finally, correlation of metabolomics data with transcriptomics (FlyAtlas.org) is very challenging due to different data formats and samples obtained at different times and from different fly tissues. A standardised method for sample collection and preparation is required to integrate tissue-specific data obtained from different omics approaches.

## 6 *In vitro* expression of *Drosophila melanogaster* CG30016 and urate oxidase

### 6.1 Introduction

Following experiments in the whole fly and individual fly tissues, *in vitro* studies on *Drosophila* CG30016 protein were performed. Metabolomics results described in the previous chapter confirmed that CG30016 is involved in purine metabolism and urate degradation pathway. However, because the work has been done in whole fly tissues and due to complexity of metabolomics data, it was impossible to observe direct effects of CG30016 knockout. In order to do so, a simple system has been employed, where the gene of interest was cloned and expressed using several different *in vitro* expression systems. Purified fly protein was then used in an enzyme assay, following assay conditions used in similar studies. The same approach has been previously described using 5-HIU hydrolase candidate proteins of other species, including mouse [102, 118], zebra fish [375] and *Escherichia coli* [377, 378]. The substrate for 5-HIU hydrolase, 5-HIU, is unstable. Its half-life at neutral pH is around 30 minutes [355]. Because of the substrate instability, the enzyme assay performed here employs two enzymatic reactions: urate conversion to 5-HIU catalysed by UO and the downstream reaction catalysed by 5-HIU hydrolase (Figure 6-1). Hence, *Drosophila melanogaster* UO and CG30016 genes were cloned and expressed, and a stable substrate, urate, was used in the assay.



**Fig. 6-1 Urate degradation pathway, enzyme assay.** Fruit fly UO and CG30016 proteins were added to urate substrate. Appearance of downstream metabolites was monitored using LC-MS or CD-spectroscopy.

#### 6.1.1 Choosing the expression system

In order to choose the most suitable expression system for the expression of fruit fly proteins, several factors were taken into consideration, including price, time, protein yield and compatibility of the protein with the system (does the system contain all



components essential for the expression of the fly gene). Three systems were compared (Table 6-1).

**Table 6-1 Comparison of three expression systems used in the study including their advantages and disadvantages**

Expression system	Advantages	Disadvantages
TNT® T7 Insect Cell Extract Protein Expression System	Expression takes only 4 hours Cost-effective Easy, quick, single-tube protocol	Low protein yield (up to 75µg/ml of protein)
Champion™ pET100 Directional TOPO® Expression Kit	High protein yield Regulated protein expression	Possibly missing <i>Drosophila</i> gene expression components
Transient expression in <i>Drosophila</i> S2 cells	Compatible with fruit fly proteins All <i>Drosophila</i> protein expression components present	Time consuming Low transfection efficiency

#### 6.1.1.1 Cell-free expression system

TNT® T7 Insect Cell Extract Protein Expression System was developed from the commonly used *Spodoptera frugiperda* Sf21 cell line [466]. It is a quick and easy protocol, which employs T7 promoter for gene expression. All components necessary for the transcription/translation are present in the TnT® T7 ICE Master Mix. To initiate protein synthesis, the only component that must be added is the DNA template. The biggest disadvantage of the system is a relatively low protein yield (up to 75µg/ml), which might be too low for the enzyme assay. Moreover, it is impossible to predict whether expressed proteins will be functional and what the effect of the transcription/translation components is going to be on the proteins.

#### 6.1.1.2 Expression using *E. coli*

*E. coli* is one of the most common systems used for protein expression. It has been previously used for the expression of proteins involved in urate degradation pathway [102, 118, 375, 377, 378]. The Champion™ pET100 Directional TOPO® Expression Kit also employs T7 RNA Polymerase and T7 promoter [467, 468]. It allows direction cloning of a

blunt-end PCR product, which is fast and does not require restriction digest and ligation reactions. Moreover, it has lac repressor gene, *lacI* and T7lac promoter, which is IPTG-inducible [469]. As a result the basal transcription from T7lac promoter is reduced and can be controlled by the addition of IPTG. This system is one of the best-studied expression systems and provides a high protein yield. However, there is a risk that the system might not have all the necessary components for the expression of a functional fruit fly protein.

#### **6.1.1.3 Drosophila S2 cells**

S2 cells were derived from primary culture *Drosophila melanogaster* embryos and have been commonly used to study *Drosophila* proteins [396, 470, 471]. The biggest advantage of this system is that it contains the complete fruit fly transcription/translation machinery. Hence, it has the best chance of expressing a functional *Drosophila* protein. The biggest disadvantage on the other hand is low transfection efficiency of S2 cells. Moreover, out of the three expression methods, this one is the most challenging and time consuming.

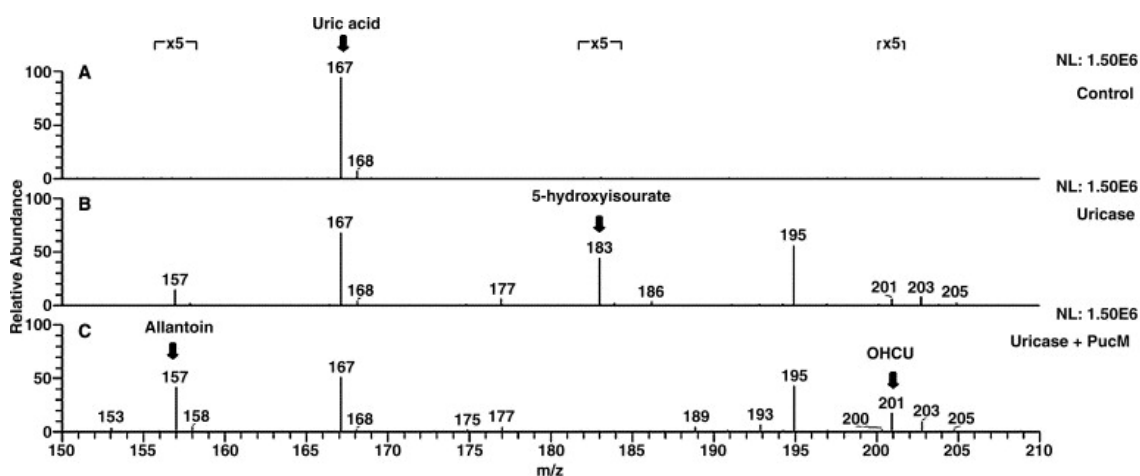
### **6.1.2 Detection of urate degradation intermediates**

In previous studies on 5-HIU hydrolase candidate proteins, various detection methods of urate degradation intermediates were used including LC-MS [102, 118, 375, 377, 378] and CD-spectroscopy [102].

#### **6.1.2.1 Detection using LC-MS**

As previously described in several studies [102, 118, 375, 377, 378], intermediates of urate degradation can be detected using LC-MS. This includes urate, 5-HIU, OHCU and allantoin. This is possible because all intermediates have different mass peaks and retention times (Figure 6-2). Moreover, it is possible to use standards to obtain 100% identification of metabolites. However, standards of unstable metabolites (5-HIU and OHCU) are not commercially available. As a result, it is only possible to run standards for urate and allantoin. This limits the reliability of this technique. Another drawback is the fact that LC-MS produces a snap shot of a metabolic state in the system, rather than real-time changes triggered by enzyme activity. In case of urate degradation pathway this is particularly challenging because the reaction occurs both spontaneously (slow reaction) and enzymatically (fast reaction). As a result, it is difficult to determine whether detected

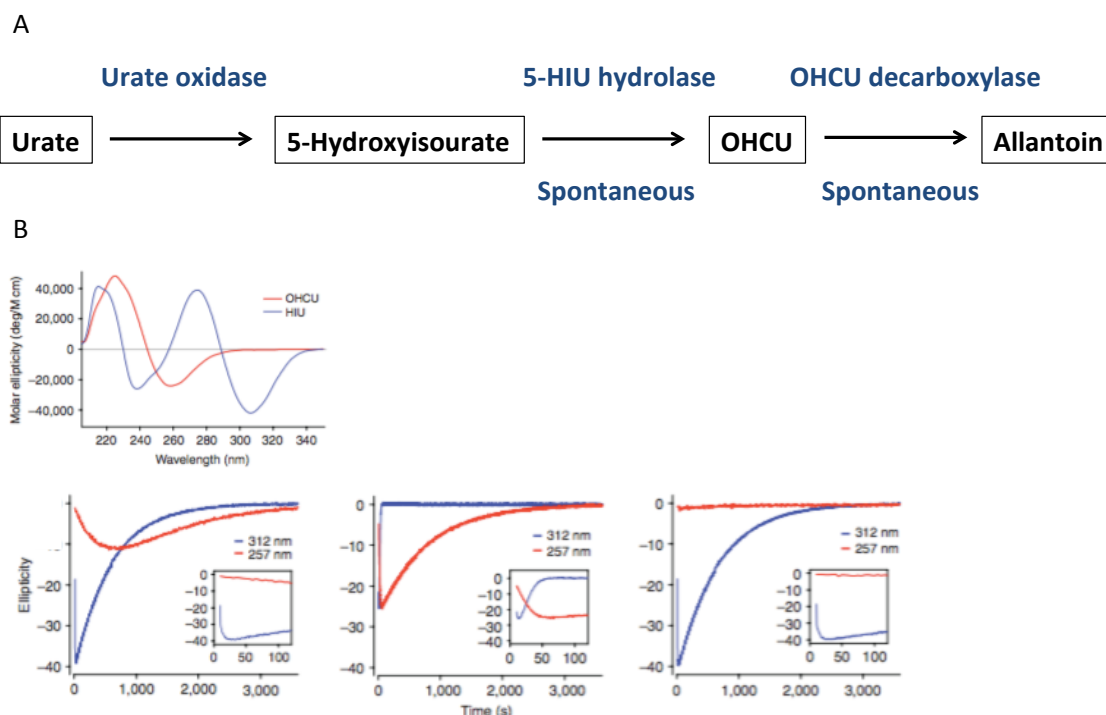
metabolites are products of spontaneous or enzymatic reaction. However, we expect to detect more products in a given time when enzymes (UO and CG30016) are added to the system.



**Fig. 6-2 LC-MS mass peaks of the urate degradation intermediates: urate, 5-HIU, OHCU and allantoin, from [384].**

#### 6.1.2.2 Detection using CD-spectroscopy

CD-spectroscopy has been previously used to detect 5-HIU and OHCU [102, 382]. It is possible because both 5-HIU and OHCU are optically active. Optically inactive urate is converted by UO to levorotatory 5-HIU, which turns into levorotatory OHCU (either spontaneously or enzymatically). Decarboxylation of OHCU (spontaneous or enzymatic) results in the production of optically inactive allantoin (Figure 6-3A). CD spectra of these compounds showed that it is possible to selectively monitor formation and decay of 5-HIU and OHCU because 5-HIU has an appreciable ellipticity at 312 nm but not 257 nm, whereas OHCU has an appreciable ellipticity at 257 nm but not 312 nm (Figure 6-3B). As a result, when UO is added to the enzyme assay, fast formation of 5-HIU can be observed followed by slow spontaneous decay accompanied by slow formation and decay of OHCU. When 5-HIU hydrolase and OHCU decarboxylase are present, the decay of 5-HIU as well as the formation and decay of OHCU are much faster. This method is particularly relevant for this assay because it monitors the speed of urate degradation and allows distinguishing the spontaneous reaction from the enzymatic one.



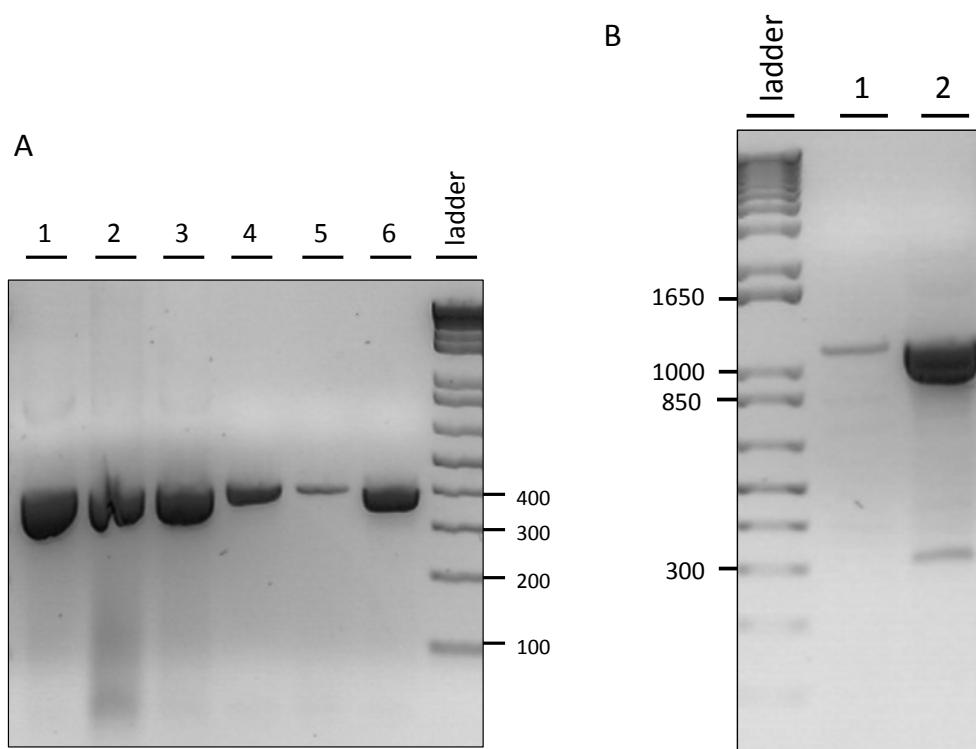
**Fig. 6-3 Urate degradation pathway monitored using CD-spectroscopy. (A) Urate degradation pathway including its intermediates and enzymatic/spontaneous reactions. (B) Approximate CD spectra of 5-HIU and OHCU (top), 5-HIU (blue) has an appreciable ellipticity at 312 nm, OHCU (red) has an appreciable ellipticity at 257 nm; (bottom) time courses of urate degradation monitored at 312 nm (blue) and 257 nm (red), (left) when UO was added, (middle) when UO and 5-HIU hydrolase was added, (right) when UO and OHCU decarboxylase was added; from [102].**

## 6.2 Results

### 6.2.1 Cell-free expression system

#### 6.2.1.1 Construct generation

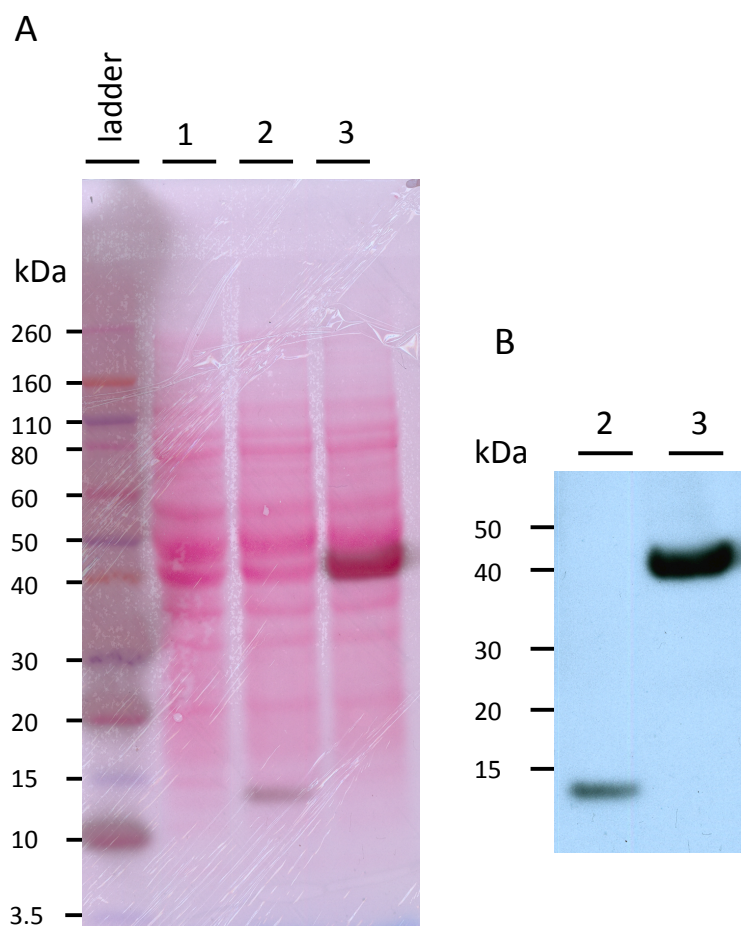
Cloning primers for *Drosophila* *UO* and *CG30016* genes were designed flanking whole gene sequences, with 6xHis tag at the C-terminal and SgfI and PmeI restriction sites. Optimal annealing temperatures were determined for both genes as described in Section 2.5.4 (see Appendix II for primer sequences): 59.2°C for *CG30016* and 63.4°C for *UO*. cDNA template for PCR was generated from RNA extracted from Canton S wild type flies. Transgenic constructs were generated as described in Section 2.8. Purified constructs were validated for the presence of genes of interest by PCR using gene-specific primers (Figure 6-4). Restriction digest was performed to confirm the orientation of inserts. Constructs were then sequenced in order to confirm that there were no mutations in cloned sequences.



**Fig. 6-4 Transgenic construct validation by PCR. (A)** PCR was performed using optimal annealing temperature and *CG30016* cloning primers. 1 – positive control 388 bp; ladder is in bp; 2-6 constructs thought to contain *CG30016* gene purified from different bacterial colonies. All constructs were validated for the presence of *CG30016* gene. **(B)** PCR was performed using optimal annealing temperature and *UO* cloning primers. 2- positive control 1108 bp; ladder in bp; 1- construct though to contain *UO* gene purified from a single bacterial colony. The construct contains the *UO* gene.

#### 6.2.1.2 Expression and protein detection

Both proteins were expressed using TNT® T7 Insect Cell Extract Protein Expression System as described in Section 2.8.3. They were then purified using HisPur Ni-NTA Spin Columns as described in Section 2.10.1. Ponceau staining and Western blotting were performed for protein detection, using anti-His primary antibody. Ponceau staining detected protein of different sizes in all samples including the negative control (Figure 6-5A). This suggests that these proteins were part of the expression system itself. Western blotting resulted in bands corresponding to expected size of *CG30016* (13.53 kDa) and *UO* (40.1 kDa) (Figure 6-5B). This confirmed that the proteins were successfully expressed using the cell-free expression system and purified thanks to the presence of 6xHis tags.



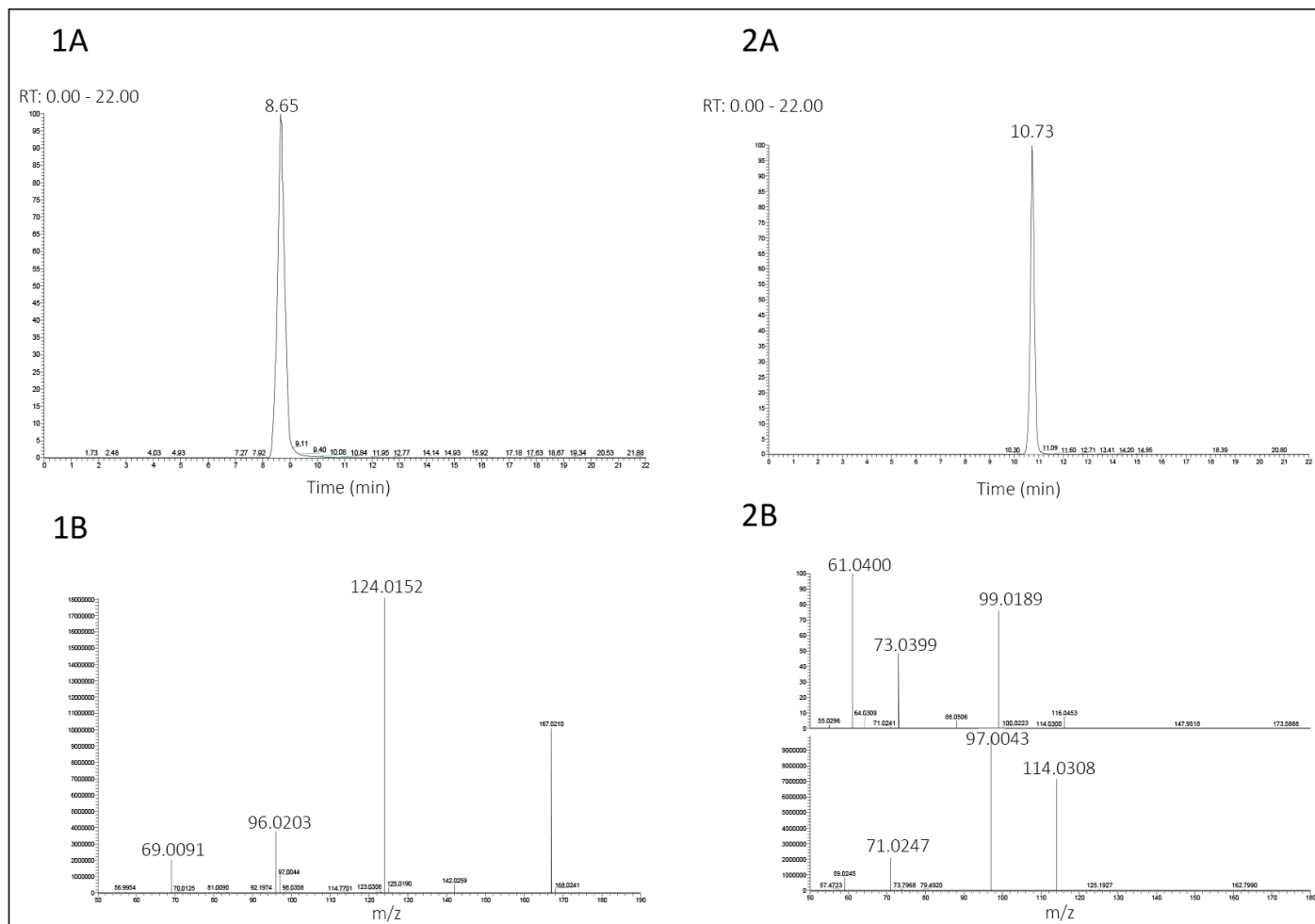
**Fig. 6-5 Detection of protein expressed using the cell-free expression system. (A) Ponceau staining and Western blotting results; 1 - negative control (empty pF25A ICE T7 Flexi® Vector), 2 – vector containing *CG30016* gene, 3 – vector containing *UO* gene. Proteins of different sizes were detected in all samples. (B) Western blotting result; 2 – *CG30016*, expected size 13.53 kDa, 3 – *UO*, 40.1 kDa. Proteins of correct sizes were detected for both *CG30016* - and *UO* – containing constructs.**

### 6.2.1.3 Enzyme assay and Liquid chromatography – mass spectrometry (LC-MS)

Purified proteins were then used for urate degradation enzyme assay (Figure 6-1) followed by LC-MS. The enzyme assay was performed using previously tested conditions [102, 118, 375, 377, 378]. Substrate solution used was 0.1 mM urate in potassium phosphate buffer, pH 7.6

For LC-MS, commercial urate and allantoin were used as standards; 100  $\mu$ M of urate and 100  $\mu$ M of allantoin. Extracted ion chromatogram for urate was at 167.02101  $m/z$  (Figure 6-6 1A), with fragmentation spectrum  $[M-H]^-$ , 167.02  $m/z$  at 8.68 min, confirming fragments of 124.01522, 96.02025 and 69.00905  $m/z$  (Figure 6-6 1B). A very intense isomer eluted at 9.3 min, resulting in a shoulder peak for urate at 8:7 min. This might be a

fragment of 5-hydroxyisourate. Allantoin eluted 10.73 min with a mass of 159.05127  $m/z$  (Figure 6-6 2A). Fragmentation spectra of allantoin were observed in both positive and negative ion mode, with slightly greater signal measured in negative ion mode: in negative ion mode  $[M-H]^-$  157.04  $m/z$  and positive  $[M-H]^+$  159.05  $m/z$  at 10.79 min. Fragments were confirmed in positive mode of 99.01887, 73.03986, 61.04004  $m/z$ , and negative mode of 114.03082, 97.00426, 71.02470  $m/z$  (Figure 6-6 2B).



**Figure 6-6 Ion chromatograms and fragmentation spectra of urate and allantoin standards. 1A Ion chromatogram for urate at 167.02101  $m/z$ , 1B fragmentation spectrum of urate at 8.68 min confirming fragments of 124.01522, 96.02025 and 69.00905  $m/z$ . 2A Ion chromatogram for allantoin at 159.05127  $m/z$ , 2B fragmentation spectrum of allantoin in negative ion mode (above) at 10.73 min confirmed fragments of 114.03082, 97.00426, 71.02470  $m/z$ , and positive ion mode (below) at 10.79 min confirmed fragments of 99.01887, 73.03986, 61.04004  $m/z$ .**



Urate, 5-hydroxyisourate, OHCU and allantoin were mass-matched. Additionally, urate and allantoin were retention time – matched and fragmentation spectrum – matched due to availability of standards (Table 6-2). Urate was detected in a negative ion mode, whereas 5-HIU, OHCU and allantoin were detected in both positive and negative modes.

**Table 6-2 Compounds detected in LC-MS: name of each compound, formula, mass as well as type of detection are provided.**

Compound Name	Formula	Monoisotopic Mass	Mass Matched	RT Matched*	Fragments Matched
Urate (Neg)	C <sub>5</sub> H <sub>4</sub> N <sub>4</sub> O <sub>3</sub>	167.02106	Yes	Yes	Yes
5-hydroxyisourate (Pos)	C <sub>5</sub> H <sub>4</sub> N <sub>4</sub> O <sub>4</sub>	185.03053	No	N/A	N/A
5-hydroxyisourate (Neg)	C <sub>5</sub> H <sub>4</sub> N <sub>4</sub> O <sub>4</sub>	183.01598	No	N/A	N/A
OCHU (Pos)	C <sub>5</sub> H <sub>5</sub> N <sub>4</sub> O <sub>5</sub>	202.03327	No	N/A	N/A
OCHU (Neg)	C <sub>5</sub> H <sub>5</sub> N <sub>4</sub> O <sub>5</sub>	200.01872	No	N/A	N/A
Allantoin (Pos)	C <sub>4</sub> H <sub>6</sub> N <sub>4</sub> O <sub>3</sub>	159.05127	Yes	Yes	Yes
Allantoin (Neg)	C <sub>4</sub> H <sub>6</sub> N <sub>4</sub> O <sub>3</sub>	157.03671	Yes	Yes	Yes

Enzyme assays were performed for a system containing urate and expressing *UO* and *CG30016* as well as several controls (Table 6-3). The controls used were:

1. A system expressing *UO* and *CG30016* without urate
2. A system expressing *UO* containing urate
3. A system expressing *UO* without urate
4. A system containing urate with an empty vector (not expressing *UO* or *CG30016*)
5. A system without urate with an empty vector (not expressing *UO* or *CG30016*)
6. A system expressing *CG30016* containing urate
7. A system expressing *CG30016* without urate

LC-MS spectra obtained for all controls containing urate (2, 4 and 6) showed the same results, i.e. urate and allantoin were detected, mass-, RT- and fragments-matched; both

5-HIU and OCHU were mass-matched. There were no differences in abundance of these compounds in different systems. The same results were obtained for the experimental sample containing urate and expressing *UO* and *CG30016*. This suggests that the presence of detected metabolites is a result of spontaneous degradation of urate, rather than the activity of expressed enzymes. Hence, it was impossible to confirm the role of *UO* and *CG30016* as urate degradation enzymes based on these results. Surprisingly, in controls without urate (1,3,5 and 7), urate was also detected and mass-, RT- and fragments-matched, so was allantoin. However, allantoin was only detected at low abundance. This could be due to equipment contamination with urate and small amounts of allantoin, perhaps from previous samples containing urate. This has to be addressed in future experiments. The reason why the enzyme assay did not produce expected metabolites might be due to poor protein yield obtained from the cell-free expression system. According to manufacturer's protocol, the system can produce 15-75 µg/ml of protein depending on type of protein being expressed. This should be sufficient for the assay. However, when Bradford assay was run, there was no difference in protein concentration between the experimental sample and the control, which only contained expression system components. This suggests that the system did not produce a lot of *UO* and *CG30016* proteins. Perhaps, this was not sufficient to catalyse urate degradation reactions. Another possibility is that the cell-free system did not contain all elements required to produce functional *UO* and *CG30016* proteins, and as a result they failed to catalyse urate degradation.

**Table 6-3 LC-MS results of enzyme assay. Several systems were used, including one expressing *UO* and *CG30016* as well as several controls. Detected compounds, their mass and type of detection are included.**

<b>Experimental: <i>UO</i> + <i>CG30016</i> + urate</b>			
<b>Compound Name</b>	<b>Mass match</b>	<b>RT match</b>	<b>Evidence</b>
Urate	Yes	Yes	RT at 8.8 min, isotope pattern, fragments match
5-hydroxyisourate	Yes	No	Mass matches at 9.3 and 11.5 min
OCHU	Yes	No	Mass matches at 4.9 and 6.4 min

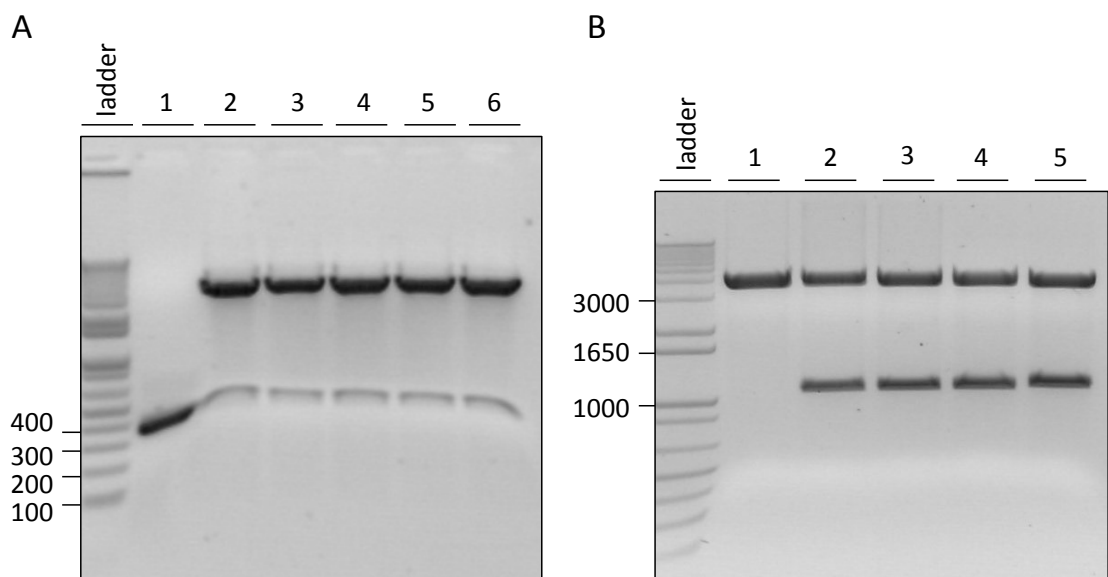
Allantoin	Yes	Yes	RT at 10.8, isotope pattern, fragments match
<b>1. UO + CG30016 – urate</b>			
Urate	Yes	Yes	RT at 8.8 min, isotope pattern, fragments match
5-hydroxyisourate	No	No	Not present
OCHU	No	No	Not Present
Allantoin	Yes	No	RT at 10.8 min (low abundance)
<b>2. UO + urate</b>			
Urate	Yes	Yes	RT at 8.75, isotope pattern, fragments match
5-hydroxyisourate	Yes	No	Mass matches at 9.2 and 11.5 min
OCHU	Yes	No	Mass matches at 4.9 and 6.6 min
Allantoin	Yes	Yes	RT at 10.8, isotope pattern, fragments match
<b>3. UO – urate</b>			
Urate	Yes	Yes	RT at 8.8 min, isotope pattern, fragments match
5-hydroxyisourate	No	No	Not present
OCHU	No	No	Not Present
Allantoin	Yes	No	RT at 10.7 min (low abundance)
<b>4. Empty vector + urate</b>			
Urate	Yes	Yes	RT at 8.8 min, isotope pattern, fragments match

5-hydroxyisourate	Yes	No	Mass matches at 9.2 and 11.3 min
OCHU	Yes	No	Mass matches at 4.9 and 6.5 min
Allantoin	Yes	Yes	RT at 10.8, isotope pattern, fragments match
<b>5. Empty vector - urate</b>			
Urate	Yes	Yes	RT at 8.7 min, isotope pattern, fragments match
5-hydroxyisourate	No	No	Not present
OCHU	No	No	Not Present
Allantoin	Yes	No	RT at 10.8 min (low abundance)
<b>6. CG30016 + urate</b>			
Urate	Yes	Yes	RT at 8.8 min, isotope pattern, fragments match
5-hydroxyisourate	Yes	No	Mass matches at 9.1 and 11.3 min
OCHU	Yes	No	Mass matches at 4.9 and 6.4 min
Allantoin	Yes	Yes	RT at 10.7, isotope pattern, fragments match
<b>7. CG30016 - urate</b>			
Urate	Yes	Yes	RT at 8.8 min, isotope pattern, fragments match
5-hydroxyisourate	No	No	Not present
OCHU	No	No	Not Present
Allantoin	Yes	No	RT at 10.8 min (low abundance)

## 6.2.2 Expression of *UO* and *CG30016* in S2 cells

### 6.2.2.1 Construct generation

Cloning primers for *UO* and *CG30016* genes were designed flanking whole gene sequences, with 6xHis tag at the C-terminal. Optimal annealing temperatures were determined for both genes as described in Section 2.4 (see Appendix II for primer sequences). cDNA template for PCR was generated from RNA extracted from Canton S wild type flies. Transgenic constructs were generated as described in Section 2.8. Purified constructs were validated for the presence of genes of interest by PCR using gene-specific primers as well as restriction digest, in order to determine the orientation of inserts (Figure 6-7). Both genes were present in transgenic constructs in the correct orientation. Finally, the constructs were sequenced in order to confirm that there were no mutations in cloned sequences.



**Fig. 6-7 Validation of transgenic constructs by restriction digest. (A) Constructs thought to contain *CG30016* gene were restriction digested EcoRI and XbaI enzymes, expected band size 435 bp was observed, confirming the presence of *CG30016* gene and its correct orientation. (B) Constructs containing *UO* gene were digested using EcoRI and XbaI, expected band sizes of 1120 bp was confirmed. *UO* gene is present in the right orientation in the transgenic construct.**

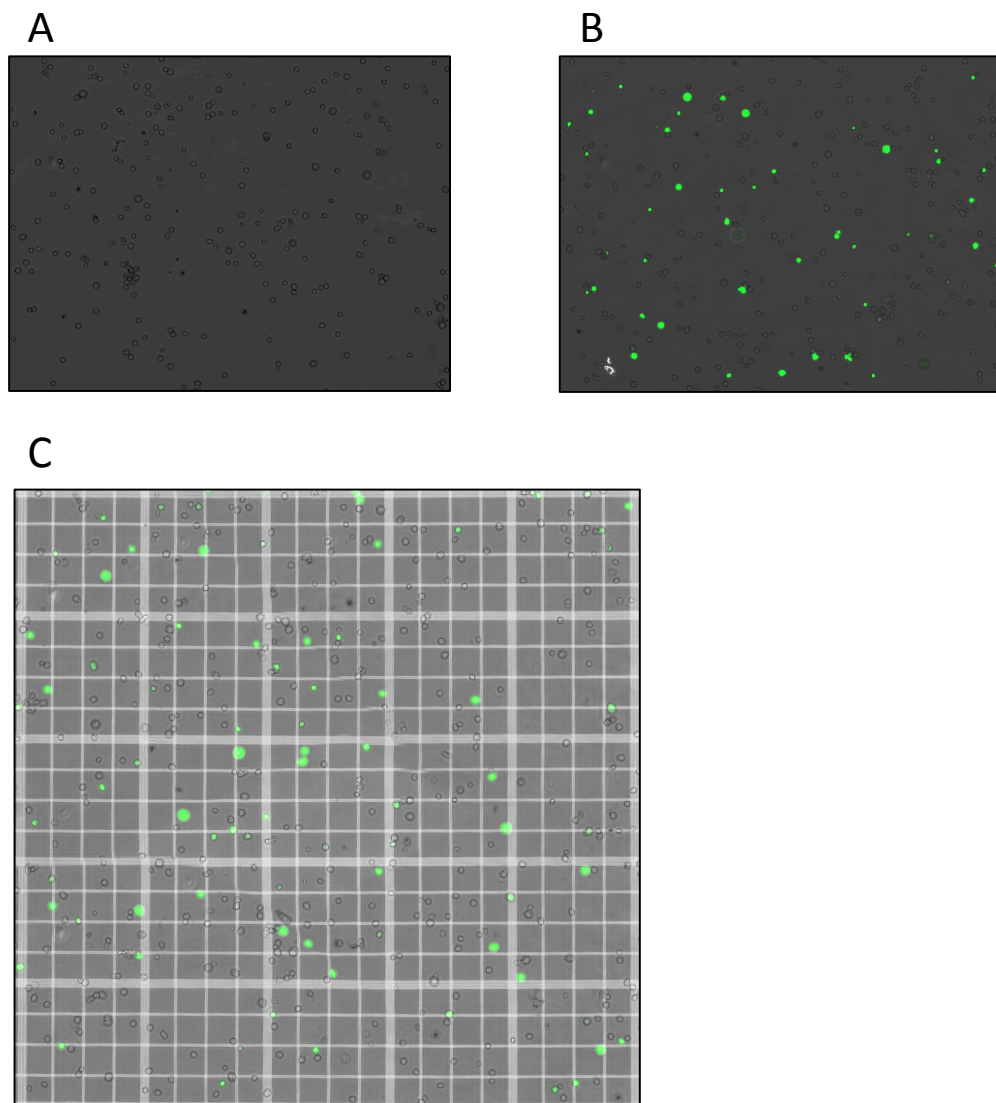
### 6.2.2.2 Transient transfection and its efficiency

Transient transfection was performed as described in Section 2.9.2. Cells were transfected or co-transfected with:

1. Empty pMT/V5-His-TOPO vector

2. pMT/V5-His-TOPO-*UO*
3. pMT/V5-His-TOPO-*UO* and pMT/V5-His-TOPO-*CG30016*
4. pMT/V5-His-TOPO-*UO*, pMT/V5-His-TOPO-*CG30016* and pMT/V5-His-TOPO-*GFP*.

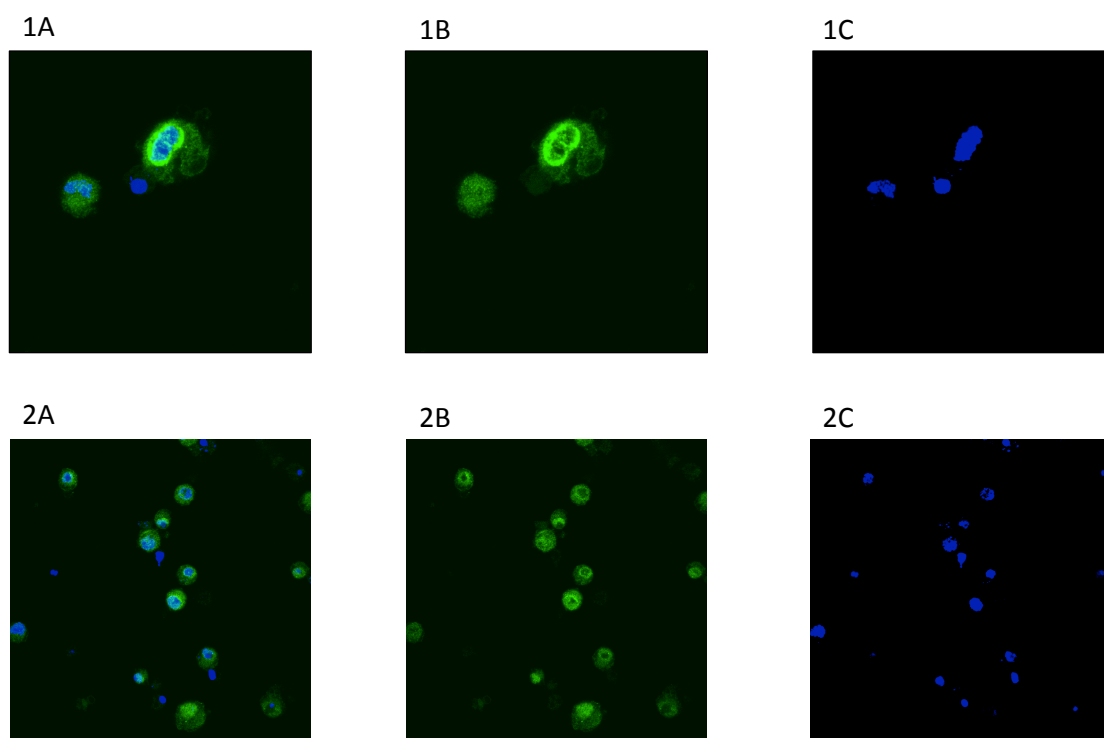
Transfection with pMT/V5-His-TOPO-*GFP* was performed in order to measure transfection efficiency. When cells are co-transfected with several constructs, it can be assumed that the cells successfully took up either all or none of the plasmids. Hence, in order to calculate transfection efficiency, cells were tested for the presence of GFP under fluorescent microscope. Sections of cell plate were examined for the presence of GFP and average transfection efficiency was calculated by measuring the ratio of GFP-positive cells versus all cells (Figure 6-8). Transfection efficiency of 18.6% was calculated. It was not clear at this stage whether this was high enough efficiency to produce high yield of *UO* and *CG30016* proteins, required for the enzyme assay.



**Fig. 6-8 Transient transfection of S2 cells. (A)** Bright-field image of transfected S2 cells showing intact 'healthy' cells. **(B)** Fluorescent microscopy image showing GFP-positive cells, which were also assumed to contain pMT/V5-His-TOPO-*UO* and pMT/V5-His-TOPO-*CG30016* constructs. **(C)** Several sections of a plate with transfected S2 cells; bright-field and fluorescent images were merged and transfection efficiency was calculated as a ratio of GFP-positive cells to all cells.

### 6.2.2.3 Detection of expressed protein

Following 48 h protein expression in S2 cells, Immunocytochemistry (ICC) assay was performed in order to confirm the expression of proteins tagged with 6xHis tag. DPAI staining was performed as well as incubation with anti-His antibody (Figure 6-9). ICC results showed that His-tagged proteins were expressed in S2 cells (Figure 6-9 1A,B and 2A,B). This suggests that both *CG30016* and *UO* were expressed in these cells.



**Fig. 6-9 Immunocytochemistry assay of S2 cells transfected with constructs containing *UO* and *CG30016* genes. (1A-C)** Cells transfected with pMT/V5-His-TOPO-*CG30016*; **(2A-C)** Cells transfected with pMT/V5-His-TOPO-*UO*; **(1A, 2A)** Image of anti-His staining (green) merged with DAPI staining (blue) showed cells expressing a protein tagged with 6xHis tag. This confirmed that *CG30016* (1A) and *UO* (2A) were expressed in these cells. **(1B, 2B)** Staining with anti-His antibody (green). **(1C, 2C)** DAPI staining nuclei of S2 cells (blue).

### 6.2.2.4 LC-MS of S2 cells

S2 cells produce urate and hence it was not necessary to add extra urate to the samples in order to observe products of urate degradation pathway. Three samples were analysed for the presence of urate, 5-HIU, OHCu and allantoin using LC-MS:

1. Empty pMT/V5-His-TOPO vector (-ve control)
2. pMT/V5-His-TOPO-*UO* (*Urate oxidase*)
3. pMT/V5-His-TOPO-*UO* and pMT/V5-His-TOPO-*CG30016* (*Urate oxidase* + *CG30016*)

The results are illustrated in Table 6-4. Urate was detected in all samples, which is consistent with the fact that S2 cells produce urate. Moreover, relative abundance of urate was slightly lower in samples 2 and 3 compared to the –ve control. This suggests that expressed *UO* might be catalysing the conversion of urate to 5-HIU in these samples. This is consistent with the detection of 5-HIU in the samples, and its higher abundance in samples 2 and 3. 5-HIU is also detected in the –ve control. This might be due to expression of *UO* in S2 cells. According to FlyAtlas, S2 cells express low levels of *UO*[4], which in this case might result in low abundance of 5-HIU in the control. OHCU was not detected in any of the samples. This is puzzling as it should be present in all samples as a result of spontaneous degradation of 5-HIU. We were also expecting to see more OHCU in sample 3 due to expression of *CG30016*. Since it was not detected at all, it was impossible to confirm whether *CG30016* was expressed as an active protein and to confirm its role as 5-HIU hydrolase. Finally, comparable abundance of allantoin was detected in all samples. This suggests that it is a product of spontaneous degradation of 5-HIU rather than an enzymatic reaction. Once again, this result did not confirm the role of *CG30016*.

**Table 6-4 LC-MS results of S2 cells transfected with three construct: empty TOPO vector (-ve control), TOPO-*UO* and TOPO-*CG30016*. Presence and relative abundance of each metabolite of urate degradation pathway is included.**

Compound name	-ve control	<i>urate oxidase</i>	<i>urate oxidase</i> + <i>CG30016</i>
Urate	Detected (36.6)	Detected (28.4)	Detected (34.3)
5-HIU	Detected (2.9)	Detected (4.1)	Detected (4.4)
OHCU	Not detected	Not detected	Not detected
Allantoin	Detected (0.7)	Detected (0.72)	Detected (0.61)

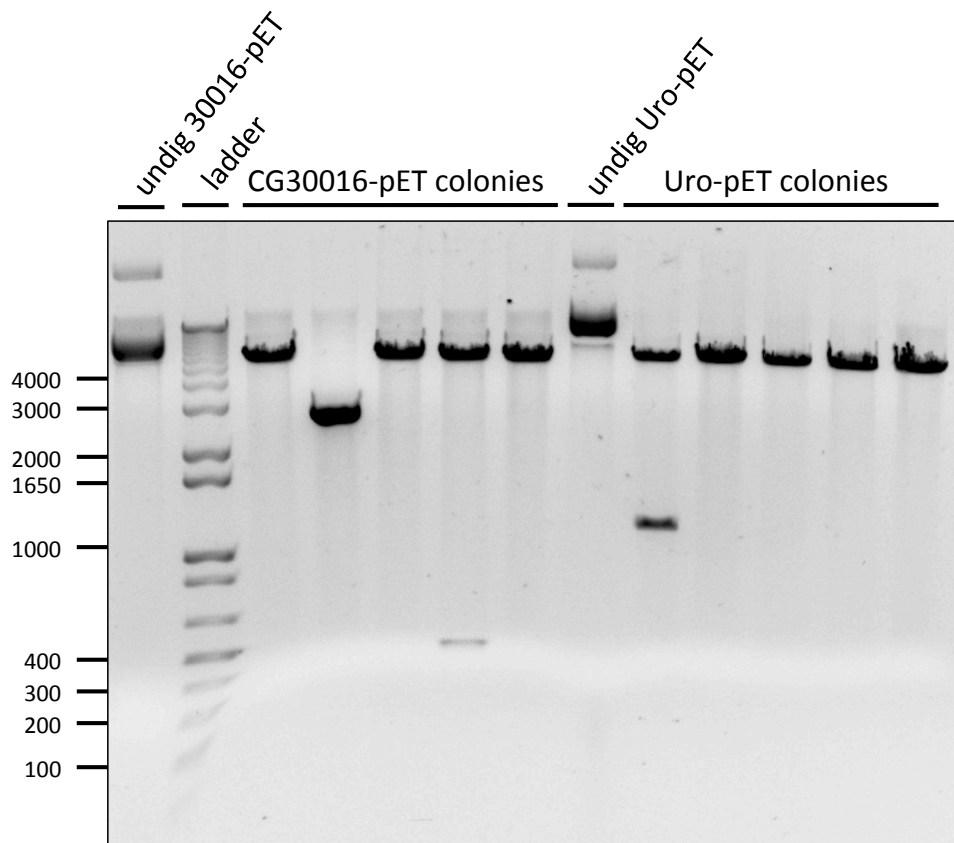
### 6.2.3 Expression of *UO* and *CG30016* in *E. coli*

#### 6.2.3.1 Construct generation: PCR, primers, gel purification, ligation

Cloning primers for *UO* and *CG30016* genes were designed flanking whole gene



sequences, with 6xHis tag at the C-terminal and EcoRI and XbaI restriction sites. Optimal annealing temperatures were determined for both genes as described in Section 2.4 (see Appendix II for primer sequences): 59.2°C for *CG30016* and 63.4°C for *UO*. cDNA template for PCR was generated from RNA extracted from Canton S wild type flies. Transgenic constructs were generated as described in Section 2.8. Purified constructs were validated for the presence of genes of interest by PCR using gene-specific primers as well as restriction digest, in order to determine the orientation of inserts. Correct orientation was confirmed for both genes (Figure 6-10). Finally, constructs validated for the presence of genes in correct orientation were sequenced and sequences without mutations were confirmed.

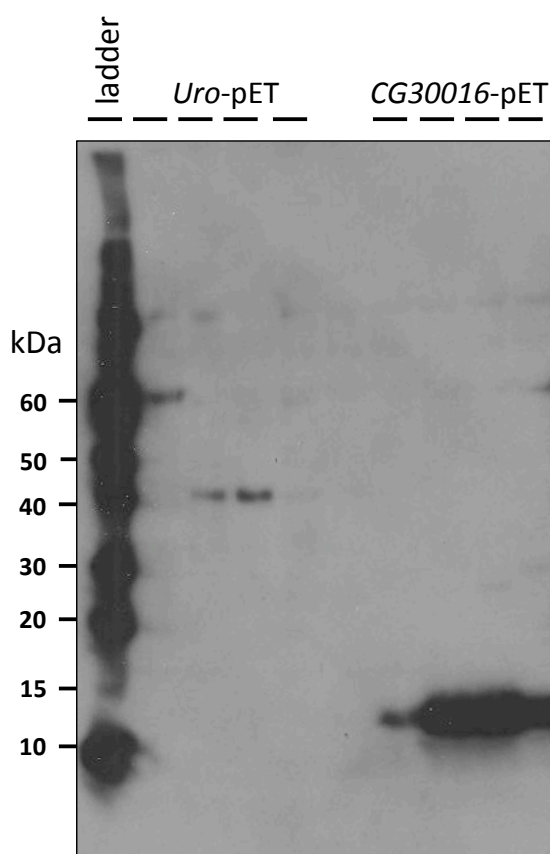


**Fig. 6-10 Validation of transgenic constructs by restriction digest.** *CG30016*-pET constructs were restriction digested using XbaI and SacI with expected band sizes of 442 bp. Correct band sizes were observed in lane 6. *UO*-pET constructs were restriction digested using XbaI and SacI with expected band size of 1159 bp. Correct band sizes were present in lane 9. These results confirmed the presence of both genes in the right orientation.

#### 6.2.3.2 Expression and protein detection

Pilot expression of confirmed constructs was performed as described in Section 2.8.10. In

order to establish optimal expression time and temperature, Western blotting was performed using anti-His antibody. The results showed that the optimal time and temperature for both constructs was 16 h induction and 37°C (Figure 6-11). Hence, scaled up reactions were performed at these conditions.



**Fig. 6-11 Determination of optimal expression time and temperature using Western blotting.** Expected band sizes are 40.1 kDa for *UO* and 13.53 kDa for *CG30016*. Lanes 2-5 are BL21 cells transfected with *UO*-pET constructs: lane 2 – uninduced BL21 cells (-ve control), lane 3 – 3 h induction at 37 °C, lane 4 – 16 h induction at 37 °C, lane 5 – 4 h induction at 28 °C. Lanes 6-9 BL21 cells transfected with *CG30016*-pET constructs: lane 6 – uninduced BL21 cells (-ve control), lane 7 - 3 h induction at 37 °C, lane 8 - 16 h induction at 37 °C, lane 9 - 4 h induction at 28 °C. Optimal temperature and time for both constructs was 16 h induction at 37 °C.

Following the expression of *UO* and *CG30016* at optimal conditions, both proteins were purified thanks to the presence of 6xHis tags. Subsequently, the protein buffer was changed to one suitable for further experiments, using Slide-A-Lyzer Dialysis Cassettes (Thermo Fisher Scientific, UK). However, protein precipitation was observed during the procedure, suggesting some protein was lost. Bradford assay was performed in order to determine the amount of expressed protein. Calculated protein yield was *CG30016*: 2.6

$\mu\text{g}/\mu\text{l}$  and  $\text{UO}$ :  $0.4 \mu\text{g}/\mu\text{l}$ .

### 6.2.3.3 Enzyme assay and CD - spectroscopy

Purified proteins were used for urate degradation enzyme assay (Figure 6-1) followed by CD-spectroscopy. In order to optimise the enzyme assay for CD-spectroscopy, the conditions were first tested using urate substrate ( $0.1 \text{ mM}$  uric acid sodium salt in potassium phosphate buffer,  $\text{pH } 7.6$ ) and commercially obtained *UO* (Sigma Aldrich, UK). Different concentrations of the substrate and enzyme were used to determine the conditions, producing the desired reaction speed for CD-spectroscopy. Initially,  $0.4 \text{ U}$  of *UO* and  $0.1 \text{ mM}$  of urate were used (Figure 6-12). In these conditions *UO* activity was fast, as the conversion had commenced within the 'dead time' of the experiment. A 20-fold dilution of the enzyme was prepared to capture the early events in the reaction.  $0.02 \text{ U}$  of *UO* and  $0.1 \text{ mM}$  of urate were used. Circular dichroism time course measurement of conversion of  $0.1 \text{ mM}$  urate was obtained at 0, 15 and 40 min (Figure 6-13A). Moreover, spectral changes over 1 h were measured (Figure 6-13B). These spectra showed an expected shift corresponding to the breakdown of urate and production and breakdown of 5-HIU and OHCU.

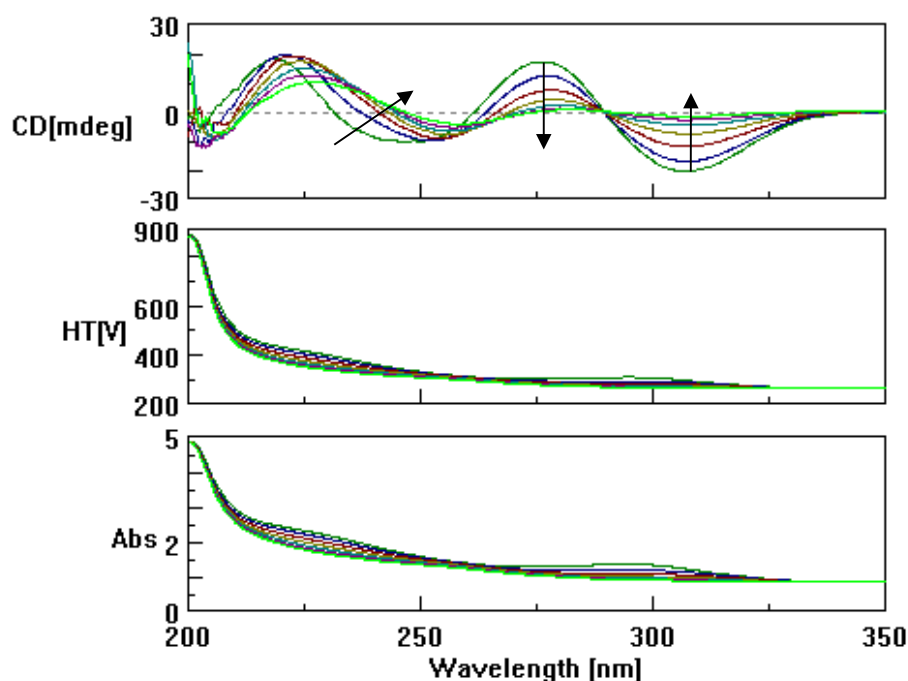
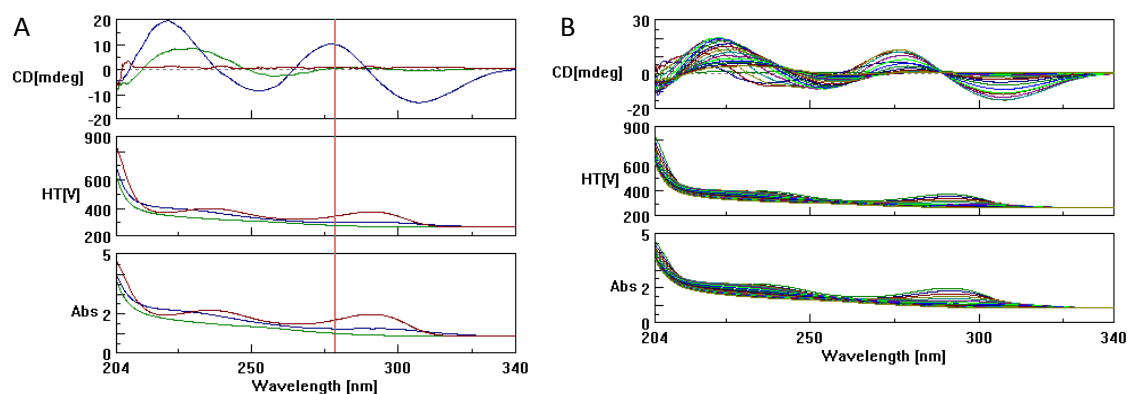
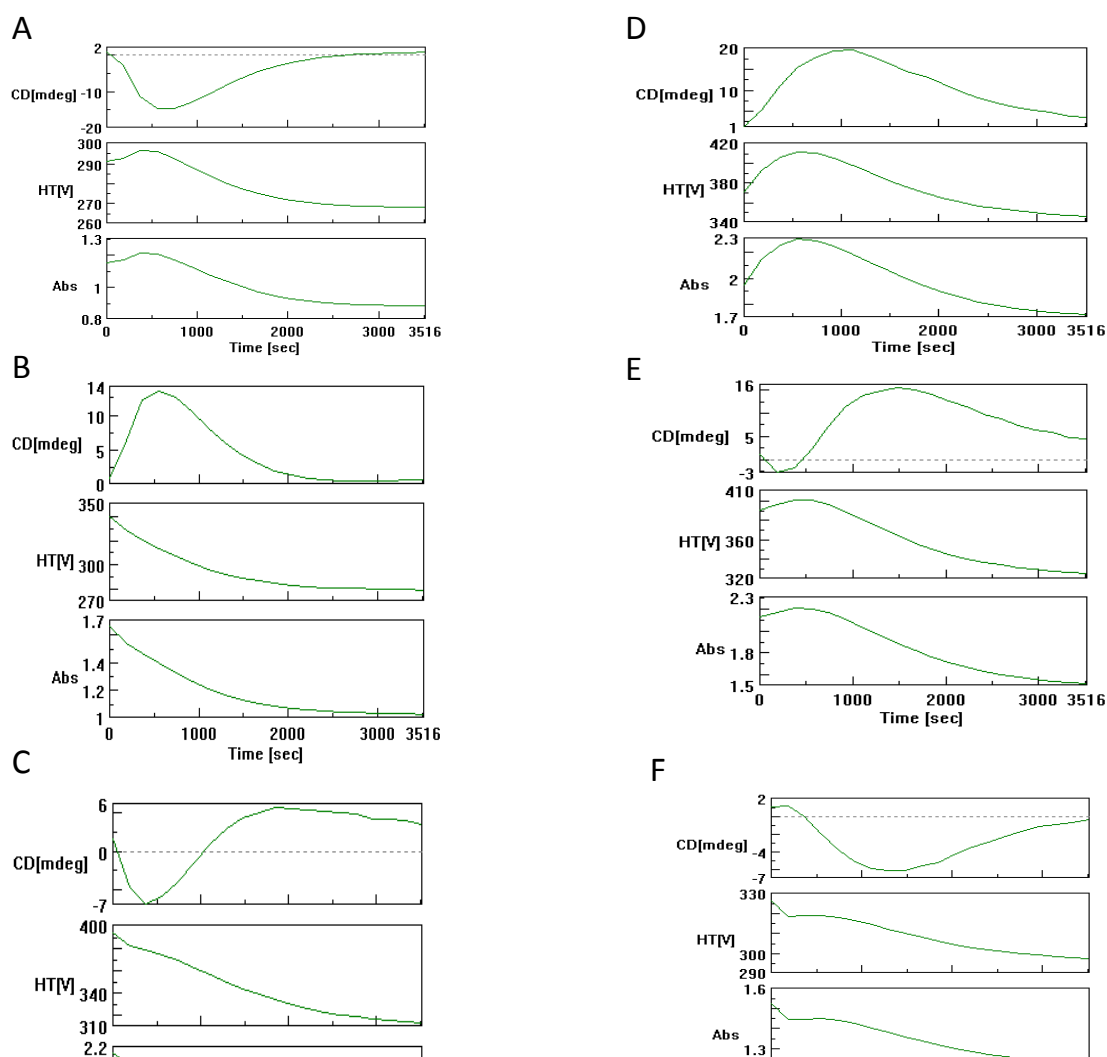


Fig. 5-12 Circular dichroism time course measurement of  $0.4 \text{ U}$  of *UO* conversion of  $0.1 \text{ mM}$  Urate (arrows denote the ellipticity trend with time). Data represent spectra obtained at  $\sim 3 \text{ min}$  intervals. *UO* activity was fast and the conversion had commenced within the 'dead time' of the experiment



**Fig. 6-13 Circular dichroism of 0.02 U of UO conversion of 0.1 mM urate. (A) CD time course measurement, data represent spectra obtained at ~0 min (brown), 15mins (blue) and 40mins (green). (B) Spectral changes over 1hr.**

Additionally, CD Ellipticity values were obtained at different wavelengths in order to confirm the production and breakdown of 5-HIU and OHCU. Values were obtained at 308 nm, 277 nm, 240 nm, 220 nm, 230 nm and 260 nm (Figure 6-14). At 308 nm, 277 nm, 240 nm and 220 nm spectra characteristic to the formation and degradation of 5-HIU were observed (Figure 6-14A-D). This corresponds to the activity of UO and the conversion of urate to 5-HIU, followed by its spontaneous degradation to OHCU. At 230 nm and 260 nm spectra characteristic to the formation of OHCU were obtained (Figure 6-14E-F). This corresponds to the spontaneous degradation of 5-HIU to OHCU.



**Fig. 6-14 CD Ellipticity values at different wavelengths following addition of 0.02 U of UO to 0.1 mM urate. (A-D) At 308 nm, 277 nm, 240 nm and 220 nm peaks corresponding to the fomration and degradation of 5-HIU were observed. (E-F) At 230 nm and 260 nm peaks characteristic to the formation of OHCU were seen.**

We expected to observe similar spectra upon the addition of expressed *Drosophila* UO and a faster conversion of 5-HIU to OHCU upon the addition of expressed *CG30016*. However, no conversion of urate to 5-HIU and OHCU was observed in these samples. This suggests that both *Drosophila* proteins were not expressed as active enzymes. Alternatively, this could be due to the yield of expressed proteins being too low or the enzyme assay conditions not suitable for this experiment. Because of this, it was impossible to determine the function of *CG30016* based on this assay.

## 6.3 Discussion

The results illustrated in this chapter were inconclusive and hence did not confirm or refute the role of *CG30016* as 5-HIU hydrolase. Unfortunately, in case of all expression systems, the *CG30016* protein was expressed at either low levels or inactive. Because of this, the enzyme assay did not produce expected results.

### 6.3.1 Problems with protein expression

In case of the cell-free expression system, the enzyme assay results suggest that detected urate degradation intermediates were produced in non-enzymatic reactions. A likely explanation for this is that *Drosophila* enzymes were expressed at low levels and hence, even if they were active, there was no noticeable difference between the experimental and controls. A possible solution would be to scale-up the expression reaction and repeat the purification step and enzyme assay followed by one of the detection methods.

The main disadvantage of S2 cells in this study was the low transfection efficiency of 18.6%. Moreover, S2 cells produce urate and express UO at low levels. This means that the cells transfected with an empty vector can still produce 5-HIU, which can then spontaneously degrade to OHCU and allantoin. As a result detected metabolites might be a result of these spontaneous reactions or the expression of *UO* and *CG30016*, and it would be impossible to distinguish between the two. However, the results indicate that the expressed UO was active, as in the cells transfected with transgenic UO, less urate and more 5-HIU was detected compared to the control. On the other hand, there was no evidence for the expression of active *CG30016* in this experiment and hence its role as 5-HIU hydrolase was not verified.

Expression in *E.coli* produced a high yield of both proteins. However, during the purification process, protein precipitation was observed. Our results suggest that the protein yield after purification was too low for the enzyme assay or that the expressed proteins were inactive. Once again, it was impossible to determine the function of *CG30016* in this experiment.

Both UO and 5-HIU hydrolase have been successfully expressed in previous studies using *E. coli* expression system [102, 118, 375, 377, 378]. It suggests that the system or the purification process require further optimisation for the expression of active *Drosophila*

*UO* and *CG30016* genes, and this should be the focus of the future work.

### **6.3.2 Metabolite detection: LC-MS versus CD-spectroscopy**

In previous studies both LC-MS and CD-spectroscopy have been used to detect metabolites of urate degradation, and to validate 5-HIU hydrolase candidate proteins in other species [102, 118, 375, 377, 378].

#### **6.3.2.1 LC-MS detection**

LC-MS has been previously used to identify mouse 5-HIU hydrolase [118]. It has been done by detecting urate, 5-HIU, OHCU and allantoin in real-time with and without 5-HIU hydrolase candidate protein. Moreover, in this study standards of all urate degradation intermediates were provided, which allowed confident identification of metabolites. In our study however, LC-MS was performed after the enzyme assay was carried out and samples were stored for several days. Even though the samples were stored at -80°C, it is possible that 5-HIU and OHCU degraded spontaneously and this interfered with ‘true’ results. Moreover, in case of S2 cells experiments, OHCU was not detected at all. This suggests that either it degraded spontaneously faster than expected, or that LC-MS did not detect it, suggesting it is not the most reliable method for the detection of urate degradation intermediates. This has been previously observed and is described in Chapter 5.

It would be interesting to perform the assay again with real-time LC-MS analysis. Moreover, it would be crucial to obtain 5-HIU and OHCU standards for confident identification of the reaction products.

#### **6.3.2.2 CD-spectroscopy detection**

CD-spectroscopy has been used to detect 5-HIU and OHCU in previous studies [102, 382]. 5-HIU and OHCU formation and decay were observed upon the addition of mouse 5-HIU hydrolase and OHCU decarboxylase candidates. Both proteins were identified using this technique. The advantage of CD-spectroscopy over LC-MS for the detection of urate degradation intermediates is the fact that CD-spectroscopy can be easily performed in real-time, without the need for standards. It is also more cost-effective.

Our results confirmed that 5-HIU and OHCU could be distinguished using CD-spectroscopy. However, due to problems with the expression of active *Drosophila*

proteins, it was impossible to use this method to determine the role of *CG30016*.

### 6.3.3 Future work

In conclusion, all expression systems did not yield enough active protein for the enzyme assay and hence it was impossible to determine the role of *CG30016* in these experiments.

In order to answer this question, the expression and purification of *CG30016* have to be optimised. This could be achieved by changing the pH, temperature and/or composition of the purification medium. Moreover, it might be useful to investigate whether changing the 6xHis tag to another tag or to the N-terminal of the protein influences its solubility and activity. The enzyme assay would then be performed and reaction products identified using CD-spectroscopy. It could be beneficial to use another identification method, such as <sup>13</sup>C NMR spectroscopy, as described in [102], in order to identify the final product of urate degradation pathway, allantoin.

Finally, it could be more feasible to perform the enzyme assay in dissected *Drosophila* tubules, and compare *CG30016* and *UO* mutant flies to a control. However, since this would use the whole tissue, the identification of metabolites using LC-MS and CD-spectroscopy might be more difficult due to many other metabolites present in the sample.



## 7 Conclusions and Future work

The primary goal of systems biology is to integrate complex omics data, and data obtained from traditional experimental studies in order to provide a holistic understanding of organismal function. One way of achieving this aim is to generate GEMs, which contain information on all metabolites, enzyme-coding genes, and biochemical reactions in a biological [304, 305]. Systems approaches [472] hold great promise in genetically tractable simple model organisms, like *Drosophila*. The fly, with its metazoan body plan and powerful genetics, provides an ideal compromise in which to work up the concept of interacting tissues in systems biology. Generation of GEMs of model organisms, including *Drosophila melanogaster*, has been recognised as a priority in systems biology [347]. It would improve our understanding of the holistic function of this species, and validate the fly as a model for studying the metabolism of other species, including humans, as well as modelling metabolic diseases, and developing and testing insecticides.

*Drosophila* GEM has been generated in our lab. However, gaps have been identified, where enzymes producing or degrading metabolites were not identified. Filling in these gaps is essential for completion of the core metabolome of the fly and its validation as a model organism. The main focus of my work was to develop a pipeline for efficient gap filling using metabolomics approaches combined with standard reverse genetics methods.

### 7.1 Development of a method for gap filling of *Drosophila* metabolic global metabolic model

I have not managed to develop a complete pipeline for metabolic gap filling. However, several approaches were examined and discussed, which pointed out necessary improvements and future work required towards effective gap filling method. Steps involved in this process were identified, including sequence analysis of the candidate gene, standard reverse genetics methods (gene knockout followed by phenotype investigation), and metabolomics analysis of the knockout. Moreover, *in vitro* experiments combined with metabolomics approaches were performed in the attempt to close the gap. All steps are described in more detail in the next section.

The main challenge in the development of gap filling method was the fact that the untargeted metabolomics approach used here is hypothesis-free in nature. As a result, it is not the best method to apply to obtain final answers to a research question. Moreover, metabolomics is a relatively young discipline in comparison to other omics, which is constantly developing in terms of techniques and data processing tools. Hence, metabolomics experiments require further optimisation and standardisation in order to be suitable for a given study.

Altogether, thanks to the work presented here, we are a step closer towards the development of a gap-filling pipeline in *Drosophila melanogaster* GEM. The areas that require further optimisation were identified and are the focus of future research.

## **7.2 CG30016 is involved in urate degradation pathway**

Even though the development of a gap-filling pipeline has not been completed in the work presented here, several important questions about the selected fruit fly candidate gene for the metabolic gap, 5-HIUH, were answered.

Based on illustrated results it is clear that *Drosophila* gene *CG30016* encodes a gene involved in purine metabolism, and specifically urate degradation pathway. These results include qPCR analysis, sequence analysis and literature review, phenotypic manifestations:

### **7.2.1 QPCR analysis**

1. A knockout of *CG30016* and *UO* genes was confirmed in mutant fly lines obtained from stock centers.
2. The most suitable larval stage, L3, was identified for further studies.
3. *UO* expression is affected and significantly reduced by the disruption of *CG30016* gene.

This correlation between expression of *UO* and *CG30016* genes confirms that the two genes are functionally related and strengthens our hypothesis that *CG30016* is involved in urate degradation pathway.

### **7.2.2 Sequence analysis and literature review**

1. Sequence analysis of *CG30016* showed that the gene has high sequence homology to 5-

HIUH of other species, and features consistent with TRPs as well as 5-HIU hydrolase.

2. A primary literature review revealed that *CG30016* is functionally-related to *UO*, another gene involved in urate degradation, due to correlated gained or loss of activity of these two genes during evolution [102].
3. FlyAtlas search showed that both *UO* and *CG30016* are enriched in Malpighian tubules (MTs), which also suggests a functional relationship [4, 6].

### **7.2.3 Phenotypic manifestations**

1. Malpighian tubules of *CG30016* mutant flies exhibit an inflated ureter phenotype.
2. There is no relationship between *CG30016* knockout and urate crystals accumulation, which revealed that the MT phenotype is not caused by urate accumulation in the ureter.
3. Fluid secretion rate by MTs is not altered in the *CG30016* knockout when compared to the control, which indicates that the observed MT phenotype is not related to aberrant fluid secretion by the tubules.
4. The phenotype could be a result of the accumulation of a metabolite associated with the *CG30016* activity, inflammation, defective osmoregulation or structural abnormality.
5. Oxidative stress survival is improved in flies without *CG30016* expression, suggesting that blocking urate degradation pathway has an antioxidative effect. This is most likely related to the accumulation of urate in *CG30016* mutants. Urate has been shown before to act as an antioxidant. However, our results suggested that the substrate of a reaction catalysed by *CG30016* was also likely to act as an antioxidant.
6. Longevity assay revealed that flies deficient for *CG30016* had a shorter lifespan than the controls when reared in normal conditions. This could be associated with pro-oxidative properties of urate accumulated in these mutants.

### **7.2.4 Metabolomics analysis**

1. The results confirmed that urate degradation is blocked in MTs of *CG30016* mutant because of build up of metabolites upstream from expected knockout and decrease of downstream metabolites.
2. However, it was not clear whether *CG30016* was the 5-HIUH. Several other hypotheses were proposed that require further investigation.
3. Metabolomics results also confirmed that tubules rather than the whole fly were more suitable for metabolomics analysis of purine metabolism.

4. Single-gene mutations not only affect levels of metabolites next to a knockout but also affect numerous unrelated metabolic pathways. This highlights how small changes affect multiple metabolic pathways in a given system.

### 7.2.5 *In vitro* experiments

Several expression systems were employed in order to express *Drosophila* *UO* and *CG30016* genes, including cell-free expression system, *E. coli*-based system, and *Drosophila* *S2* cells. None of the systems resulted in high levels of expression of active *CG30016* protein. The role of *CG30016* gene has not been established in these experiments. However, properties of *CG30016* protein were observed:

1. *UO* and *CG30016* were expressed using cell-free expression system at low levels and hence it was not possible to detect the products of enzymatic reactions of urate degradation in LC-MS measurements.
2. *UO* and *CG30016* were expressed in *Drosophila* *S2* cells. However, the transfection efficiency was too low to observe the products of enzymatic reactions of urate degradation using LC-MS.
3. *UO* and *CG30016* were expressed in *E. coli* at high levels. However, problems with protein precipitation resulted in the low final yield of both proteins, and/or proteins being inactive. Hence, it was impossible to determine the role of *CG30016* in this experiment.
4. LC-MS is not the most suitable method for the detection of urate degradation intermediates following the enzyme assay because of poor reproducibility of LC-MS measurements and the instability of 5-HIU and OHCU.
5. CD-spectroscopy was validated as a method for detection of intermediates of urate degradation, and distinguishing between 5-HIU and OHCU.

### 7.2.6 Future work

#### 1. Mutant phenotypes:

- In order to understand the molecular basis of the inflated ureter phenotype further studies should be conducted, for example in-depth morphological studies of the tubules.
- Inflated abdomen in the mutant flies could be a result of increased haemolymph water content. In order to test this hypothesis, further experiments are required, for example wet-dry weight measurement (as described in [69]).

#### 2. Metabolomics analysis:

- Metabolomics results revealed the need to obtain authentic standards for the accurate metabolite identification. This is particularly challenging in the case of 5-HIU and OHCU, which are unstable and not available from commercial sources. A possible solution would

be to synthesise these compounds using commercial enzymes involved in urate degradation pathway, followed by optimising a storage method for both compounds in order to increase their stability.

- It is of crucial importance to distinguish between the spontaneous and enzymatic urate degradation pathway. One of the methods would be to differentiate between S- and R-allantoin, for example <sup>13</sup>C NMR spectroscopy, as described in [102].
- Performing the fluid secretion assay in *CG30016* mutant and the control followed by LC-MS analysis should be performed as the detection of metabolites of interest was more robust in these samples in comparison to whole fly or individual tissues.

### 3. *In vitro* expression:

- Optimisation of in vitro expression of *UO* and *CG30016* is required in order to obtain a high yield of active protein. Optimisation should focus on protein purification method, including pH, temperature and/or composition of the purification medium adjustment.
- Moreover, a possible method for protein expression at higher yield would be to generate a stable S2 cell line expressing both genes (as described in [473]).
- Another detection method should be used alongside CD-spectroscopy. For example in order to reduce spontaneous degradation of purine metabolism intermediates, real-time LC-MS could be performed following the enzyme assay.
- Finally, enzyme assay could be performed in dissected MTs (*CG30016* and *UO* mutant MTs compared to the control), followed by metabolite identification using LC-MS or CD-spectroscopy.

## 7.3 Different *Drosophila* tissues have different metabolic profiles

Base on the work on tissue-specific metabolomes presented in this thesis, and the comparison of my results with the results obtained by collaborators, several conclusions can be drawn:

1. Tissue metabolomes differ significantly from each other and from the whole fly.
2. Some of these differences can be correlated to the tissue function.
3. Some metabolites and metabolic pathways are conserved among tissues.
4. These results are consistent with FlyAtlas findings that tissue transcriptomes differ and the differences are functionally significant [4].
5. The results illustrate the need to study individual tissues as well as the whole organism. It

is clear that some metabolites that play an important role in a given tissue might not be detected in the whole fly sample because their abundance is much lower in comparison to other metabolites present in all tissues, which prevent the detection of the tissue-specific compound.

6. Isolated Malpighian tubules are suitable for the study of purine metabolism.
7. LC-MS is a suitable method for metabolomics analysis of *Drosophila* whole fly as well as individual tissues.
8. Statistical analysis, data interpretation and visualisation represent the major challenge in metabolomics studies of individual tissues and require further optimisation.

### **7.3.1 LC-MS measurement of secreted fluid**

Methodology for the study of secreted fluid has been presented in this thesis for the first time. It included sample preparation, LC-MS metabolomics and data processing. The results showed that LC-MS is a suitable method for metabolomics analysis of secreted fluid. Moreover, data analysis revealed that compounds involved in purine metabolism are detected in the secreted fluid in high abundance. Hence, this method could be utilised to study purine metabolism mutants.

### **7.3.2 Future work**

1. In order to generate a comprehensive atlas of tissue-specific metabolomes, different fly tissues have to be carefully selected in terms of their function. For example, CNS sample will provide more informative results than the head, which houses functionally diverse tissues.
2. Tissues should be selected that correspond to transcriptomics data. Metabolomics analysis should be run in parallel to transcriptomics analysis.
3. Due to high variability between replicates, more replicates of each sample should be analysed, as many as is feasible (>5).
4. Several metabolomics platforms can be combined in order to increase metabolite coverage
5. Constant efforts should be made to provide the most suitable statistical analysis of these complex datasets. Moreover, data interpretation and visualisation could be improved. This requires expertise in available bioinformatics, which is continuously expanding.

Altogether LC-MS validated as a method suitable for the study of *Drosophila* whole fly, individual tissues, as well as secreted fluid. Moreover, preliminary studies carried out by myself (not presented in this thesis) showed that the method is also suitable for the study of *Drosophila* haemolymph. LC-MS is a leading method in untargeted metabolomics and the results presented here validate it as a suitable tool for studying fruit fly specimen.

## 7.4 Final conclusion

*Drosophila melanogaster* GEM generated in our lab has many metabolic gaps. The majority of the work presented in this thesis was focused on the development of a pipeline for efficient gap filling using metabolomics methods. The process of the pipeline generation is described and discussed, identifying limitations and future work.

The work presented in this thesis provided a characterisation of the *Drosophila CG30016* gene. It revealed that the gene is involved in purine metabolism, and specifically urate degradation pathway. Moreover, characteristic phenotypes were identified as a result of *CG30016* knockout that might reflect kidney disorders in hyperuricemia in humans. If this hypothesis is confirmed, a fly model of hyperuricemia could be developed in order to test new therapies for the disease.

Moreover, methods for the development of tissue-specific metabolic maps have been evaluated and future work identified. It is clear from this study that it is important to study individual tissues in the context of the whole organism in order to understand holistic organismal functions as well as study individual genes, proteins and pathways.

Future efforts should be focused on the standardisation of metabolomics methods for *Drosophila melanogaster* research as well as metabolomics data processing, interpretation and visualisation. Moreover, further curating of *Drosophila* GEM is required. The priority is to establish biophysical constraints on reaction and transport rates once these are available from experimental data [335]. Once a constraint-based model is available, it can be used to run simulations and test hypotheses, for example, the effects of a single-gene knockout on the metabolism, and as a result validate *Drosophila melanogaster* as a model organism for studying human metabolism and metabolic disorders, vector insects and novel pesticides, and poorly characterised genes, including



global orphan enzymes.

## 8 References

1. Yamaguchi, N., et al., *Screening of SLC25A13 mutations in early and late onset patients with citrin deficiency and in the Japanese population: Identification of two novel mutations and establishment of multiple DNA diagnosis methods for nine mutations*. Hum Mutat, 2002. **19**(2): p. 122-30.
2. Adams, M.D., et al., *The genome sequence of Drosophila melanogaster*. Science, 2000. **287**(5461): p. 2185-95.
3. Millburn, G.H., et al., *FlyBase portals to human disease research using Drosophila models*. Dis Model Mech, 2016. **9**(3): p. 245-52.
4. Robinson, S.W., et al., *FlyAtlas: database of gene expression in the tissues of Drosophila melanogaster*. Nucleic Acids Res, 2013. **41**(Database issue): p. D744-50.
5. Chien, S., et al., *Homophila: human disease gene cognates in Drosophila*. Nucleic Acids Res, 2002. **30**(1): p. 149-51.
6. Chintapalli, V.R., J. Wang, and J.A. Dow, *Using FlyAtlas to identify better Drosophila melanogaster models of human disease*. Nat Genet, 2007. **39**(6): p. 715-20.
7. Rubin, G.M., et al., *Comparative genomics of the eukaryotes*. Science, 2000. **287**(5461): p. 2204-15.
8. Yamamoto, S., et al., *A drosophila genetic resource of mutants to study mechanisms underlying human genetic diseases*. Cell, 2014. **159**(1): p. 200-14.
9. Dow, J.A., *The Drosophila phenotype gap - and how to close it*. Brief Funct Genomic Proteomic, 2003. **2**(2): p. 121-7.
10. Brown, S.D. and J. Peters, *Combining mutagenesis and genomics in the mouse--closing the phenotype gap*. Trends Genet, 1996. **12**(11): p. 433-5.
11. Dietzl, G., et al., *A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila*. Nature, 2007. **448**(7150): p. 151-6.
12. Rubin, G.M. and A.C. Spradling, *Genetic transformation of Drosophila with transposable element vectors*. Science, 1982. **218**(4570): p. 348-53.
13. Bender, W., et al., *Molecular Genetics of the Bithorax Complex in Drosophila melanogaster*. Science, 1983. **221**(4605): p. 23-9.
14. Campuzano, S., et al., *Molecular genetics of the achaete-scute gene complex of D. melanogaster*. Cell, 1985. **40**(2): p. 327-38.
15. Kidd, S., T.J. Lockett, and M.W. Young, *The Notch locus of Drosophila melanogaster*. Cell, 1983. **34**(2): p. 421-33.
16. Zachar, Z. and P.M. Bingham, *Regulation of white locus expression: the structure of mutant alleles at the white locus of Drosophila melanogaster*. Cell, 1982. **30**(2): p. 529-41.
17. Rubin, G.M., *Mobile Genetic Elements* ed. J. Shapiro. 1983, New York
18. O'Hare, K. and G.M. Rubin, *Structures of P transposable elements and their sites of insertion and excision in the Drosophila melanogaster genome*. Cell, 1983. **34**(1): p. 25-35.
19. Laski, F.A., D.C. Rio, and G.M. Rubin, *Tissue specificity of Drosophila P element transposition is regulated at the level of mRNA splicing*. Cell, 1986. **44**(1): p. 7-19.
20. Rio, D.C., F.A. Laski, and G.M. Rubin, *Identification and immunochemical analysis of biologically active Drosophila P element transposase*. Cell, 1986. **44**(1): p. 21-32.
21. O'Kane, C.J. and W.J. Gehring, *Detection in situ of genomic regulatory elements in Drosophila*. Proc Natl Acad Sci U S A, 1987. **84**(24): p. 9123-7.
22. Geyer, P.K., et al., *Genetic instability in Drosophila melanogaster: P-element*

- mutagenesis by gene conversion*. Proc Natl Acad Sci U S A, 1988. **85**(17): p. 6455-9.
23. Sentry, J.W. and K. Kaiser, *Application of inverse PCR to site-selected mutagenesis of Drosophila*. Nucleic Acids Res, 1994. **22**(16): p. 3429-30.
  24. Bellen, H.J., et al., *The BDGP gene disruption project: single transposon insertions associated with 40% of Drosophila genes*. Genetics, 2004. **167**(2): p. 761-81.
  25. Spradling, A.C., et al., *The Berkeley Drosophila Genome Project gene disruption project: Single P-element insertions mutating 25% of vital Drosophila genes*. Genetics, 1999. **153**(1): p. 135-77.
  26. Thibault, S.T., et al., *A complementary transposon tool kit for Drosophila melanogaster using P and piggyBac*. Nat Genet, 2004. **36**(3): p. 283-7.
  27. Handler, A.M. and R.A. Harrell, 2nd, *Germline transformation of Drosophila melanogaster with the piggyBac transposon vector*. Insect Mol Biol, 1999. **8**(4): p. 449-57.
  28. Hacker, U., et al., *piggyBac-based insertional mutagenesis in the presence of stably integrated P elements in Drosophila*. Proc Natl Acad Sci U S A, 2003. **100**(13): p. 7720-5.
  29. Horn, C., et al., *piggyBac-based insertional mutagenesis and enhancer detection as a tool for functional insect genomics*. Genetics, 2003. **163**(2): p. 647-61.
  30. Metaxakis, A., et al., *Minos as a genetic and genomic tool in Drosophila melanogaster*. Genetics, 2005. **171**(2): p. 571-81.
  31. Franz, G. and C. Savakis, *Minos, a new transposable element from Drosophila hydei, is a member of the Tc1-like family of transposons*. Nucleic Acids Res, 1991. **19**(23): p. 6646.
  32. Brand, A.H. and N. Perrimon, *Targeted gene expression as a means of altering cell fates and generating dominant phenotypes*. Development, 1993. **118**(2): p. 401-15.
  33. Fischer, J.A., et al., *GAL4 activates transcription in Drosophila*. Nature, 1988. **332**(6167): p. 853-6.
  34. Webster, N., et al., *The yeast UASG is a transcriptional enhancer in human HeLa cells in the presence of the GAL4 trans-activator*. Cell, 1988. **52**(2): p. 169-78.
  35. Mohr, S.E., et al., *RNAi screening comes of age: improved techniques and complementary approaches*. Nat Rev Mol Cell Biol, 2014. **15**(9): p. 591-600.
  36. Dow, J.A., *Integrative physiology, functional genomics and the phenotype gap: a guide for comparative physiologists*. J Exp Biol, 2007. **210**(Pt 9): p. 1632-40.
  37. Ejsmont, R.K. and B.A. Hassan, *The Little Fly that Could: Wizardry and Artistry of Drosophila Genomics*. Genes (Basel), 2014. **5**(2): p. 385-414.
  38. Ugur, B., K. Chen, and H.J. Bellen, *Drosophila tools and assays for the study of human diseases*. Dis Model Mech, 2016. **9**(3): p. 235-44.
  39. Wangler, M.F., S. Yamamoto, and H.J. Bellen, *Fruit flies in biomedical research*. Genetics, 2015. **199**(3): p. 639-53.
  40. Mohr, S.E., et al., *Resources for functional genomics studies in Drosophila melanogaster*. Genetics, 2014. **197**(1): p. 1-18.
  41. Stapleton, M., et al., *A Drosophila full-length cDNA resource*. Genome Biol, 2002. **3**(12): p. RESEARCH0080.
  42. Bellen, H.J., et al., *The Drosophila gene disruption project: progress using transposons with distinctive site specificities*. Genetics, 2011. **188**(3): p. 731-43.
  43. Boutros, M., et al., *Genome-wide RNAi analysis of growth and viability in Drosophila cells*. Science, 2004. **303**(5659): p. 832-5.
  44. Flockhart, I.T., et al., *FlyRNAi.org--the database of the Drosophila RNAi screening center: 2012 update*. Nucleic Acids Res, 2012. **40**(Database issue): p. D715-9.

45. Bellen, H.J., C. Tong, and H. Tsuda, *100 years of Drosophila research and its impact on vertebrate neuroscience: a history lesson for the future*. Nat Rev Neurosci, 2010. **11**(7): p. 514-22.
46. Morgan, T.H., *Localization of the Hereditary Material in the Germ Cells*. Proc Natl Acad Sci U S A, 1915. **1**(7): p. 420-9.
47. Sturtevant, A.H., C.B. Bridges, and T.H. Morgan, *The Spatial Relations of Genes*. Proc Natl Acad Sci U S A, 1919. **5**(5): p. 168-73.
48. Muller, H.J., *The Production of Mutations by X-Rays*. Proc Natl Acad Sci U S A, 1928. **14**(9): p. 714-26.
49. Sturtevant, A.H. and T.H. Morgan, *Reverse Mutation of the Bar Gene Correlated with Crossing Over*. Science, 1923. **57**(1487): p. 746-7.
50. Stankiewicz, P. and J.R. Lupski, *The genomic basis of disease, mechanisms and assays for genomic disorders*. Genome Dyn, 2006. **1**: p. 1-16.
51. Kleinjan, D.A. and V. van Heyningen, *Long-range control of gene expression: emerging mechanisms and disruption in disease*. Am J Hum Genet, 2005. **76**(1): p. 8-32.
52. Sturtevant, A.H., *The Effects of Unequal Crossing over at the Bar Locus in Drosophila*. Genetics, 1925. **10**(2): p. 117-47.
53. Poulson, D.F., *Chromosomal Deficiencies and the Embryonic Development of Drosophila Melanogaster*. Proc Natl Acad Sci U S A, 1937. **23**(3): p. 133-7.
54. Louvi, A. and S. Artavanis-Tsakonas, *Notch and disease: a growing field*. Semin Cell Dev Biol, 2012. **23**(4): p. 473-80.
55. Lewis, E.B., *A gene complex controlling segmentation in Drosophila*. Nature, 1978. **276**(5688): p. 565-70.
56. Krumlauf, R., *Hox genes in vertebrate development*. Cell, 1994. **78**(2): p. 191-201.
57. Shah, N. and S. Sukumar, *The Hox genes and their roles in oncogenesis*. Nat Rev Cancer, 2010. **10**(5): p. 361-71.
58. Nusslein-Volhard, C. and E. Wieschaus, *Mutations affecting segment number and polarity in Drosophila*. Nature, 1980. **287**(5785): p. 795-801.
59. Leyssen, M. and B.A. Hassan, *A fruitfly's guide to keeping the brain wired*. EMBO Rep, 2007. **8**(1): p. 46-50.
60. Bodmer, R., *The gene tinman is required for specification of the heart and visceral muscles in Drosophila*. Development, 1993. **118**(3): p. 719-29.
61. Rieder, L.E. and E.N. Larschan, *Wisdom from the fly*. Trends Genet, 2014. **30**(11): p. 479-81.
62. Kamleh, M.A., et al., *Metabolomic profiling of Drosophila using liquid chromatography Fourier transform mass spectrometry*. FEBS Lett, 2008. **582**(19): p. 2916-22.
63. Wang, J., et al., *Function-informed transcriptome analysis of Drosophila renal tubule*. Genome Biol, 2004. **5**(9): p. R69.
64. Feany, M.B. and W.W. Bender, *A Drosophila model of Parkinson's disease*. Nature, 2000. **404**(6776): p. 394-8.
65. Johnson, K., E. Knust, and H. Skaer, *bloated tubules (blot) encodes a Drosophila member of the neurotransmitter transporter family required for organisation of the apical cytocortex*. Dev Biol, 1999. **212**(2): p. 440-54.
66. Jack, J. and G. Myette, *Mutations that alter the morphology of the malpighian tubules in Drosophila*. Dev Genes Evol, 1999. **209**(9): p. 546-54.
67. Dow, J.A., *The multifunctional Drosophila melanogaster V-ATPase is encoded by a multigene family*. J Bioenerg Biomembr, 1999. **31**(1): p. 75-83.
68. Allan, A.K., et al., *Genome-wide survey of V-ATPase genes in Drosophila reveals a*

- conserved renal phenotype for lethal alleles*. *Physiol Genomics*, 2005. **22**(2): p. 128-38.
69. Cabrero, P., et al., *Chloride channels in stellate cells are essential for uniquely high secretion rates in neuropeptide-stimulated Drosophila diuresis*. *Proc Natl Acad Sci U S A*, 2014. **111**(39): p. 14301-6.
  70. Melnick, M.B., E. Noll, and N. Perrimon, *The Drosophila stubarista phenotype is associated with a dosage effect of the putative ribosome-associated protein D-p40 on spineless*. *Genetics*, 1993. **135**(2): p. 553-64.
  71. Broderick, K.E., et al., *Interactions between epithelial nitric oxide signaling and phosphodiesterase activity in Drosophila*. *Am J Physiol Cell Physiol*, 2003. **285**(5): p. C1207-18.
  72. Broderick, K.E., et al., *Ectopic expression of bovine type 5 phosphodiesterase confers a renal phenotype in Drosophila*. *J Biol Chem*, 2004. **279**(9): p. 8159-68.
  73. Cabrero, P., et al., *A conserved domain of alkaline phosphatase expression in the Malpighian tubules of dipteran insects*. *J Exp Biol*, 2004. **207**(Pt 19): p. 3299-305.
  74. Day, J.P., et al., *Identification of two partners from the bacterial Kef exchanger family for the apical plasma membrane V-ATPase of Metazoa*. *J Cell Sci*, 2008. **121**(Pt 15): p. 2612-9.
  75. Chintapalli, V.R., et al., *Transport proteins NHA1 and NHA2 are essential for survival, but have distinct transport modalities*. *Proc Natl Acad Sci U S A*, 2015. **112**(37): p. 11720-5.
  76. Hilliker, A.J., et al., *Urate-null rosy mutants of Drosophila melanogaster are hypersensitive to oxygen stress*. *Proc Natl Acad Sci U S A*, 1992. **89**(10): p. 4343-7.
  77. Jackson, G.R., et al., *Polyglutamine-expanded human huntingtin transgenes induce degeneration of Drosophila photoreceptor neurons*. *Neuron*, 1998. **21**(3): p. 633-42.
  78. Jaiswal, M., et al., *Probing mechanisms that underlie human neurodegenerative diseases in Drosophila*. *Annu Rev Genet*, 2012. **46**: p. 371-96.
  79. Casci, I. and U.B. Pandey, *A fruitful endeavor: modeling ALS in the fruit fly*. *Brain Res*, 2015. **1607**: p. 47-74.
  80. Sehgal, A. and E. Mignot, *Genetics of sleep and sleep disorders*. *Cell*, 2011. **146**(2): p. 194-207.
  81. Schott, J.J., et al., *Congenital heart disease caused by mutations in the transcription factor NKX2-5*. *Science*, 1998. **281**(5373): p. 108-11.
  82. Wolf, M.J. and H.A. Rockman, *Drosophila, genetic screens, and cardiac function*. *Circ Res*, 2011. **109**(7): p. 794-806.
  83. Smith, W.W., et al., *From fat fruit fly to human obesity*. *Physiol Behav*, 2014. **136**: p. 15-21.
  84. Kaun, K.R., A.V. Devineni, and U. Heberlein, *Drosophila melanogaster as a model to study drug addiction*. *Hum Genet*, 2012. **131**(6): p. 959-75.
  85. Rajan, A. and N. Perrimon, *Of flies and men: insights on organismal metabolism from fruit flies*. *BMC Biol*, 2013. **11**: p. 38.
  86. Owusu-Ansah, E. and N. Perrimon, *Modeling metabolic homeostasis and nutrient sensing in Drosophila: implications for aging and metabolic diseases*. *Dis Model Mech*, 2014. **7**(3): p. 343-50.
  87. Hill, C.A., et al., *Arthropod-borne diseases: vector control in the genomics era*. *Nat Rev Microbiol*, 2005. **3**(3): p. 262-8.
  88. Hemingway, J., et al., *The molecular basis of insecticide resistance in mosquitoes*. *Insect Biochem Mol Biol*, 2004. **34**(7): p. 653-65.
  89. Raymond-Delpech, V., et al., *Ion channels: molecular targets of neuroactive*

- insecticides*. Invert Neurosci, 2005. **5**(3-4): p. 119-33.
90. Brinzer, R.A., et al., *Metabolomic profiling of permethrin-treated Drosophila melanogaster identifies a role for tryptophan catabolism in insecticide survival*. Insect Biochem Mol Biol, 2015. **67**: p. 74-86.
  91. Pandey, U.B. and C.D. Nichols, *Human disease models in Drosophila melanogaster and the role of the fly in therapeutic drug discovery*. Pharmacol Rev, 2011. **63**(2): p. 411-36.
  92. Jin, P., et al., *Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of fragile X tremor/ataxia syndrome*. Neuron, 2007. **55**(4): p. 556-64.
  93. Yang, L., et al., *Fragile X mental retardation protein modulates the fate of germline stem cells in Drosophila*. Hum Mol Genet, 2007. **16**(15): p. 1814-20.
  94. Chang, S., et al., *Identification of small molecules rescuing fragile X syndrome phenotypes in Drosophila*. Nat Chem Biol, 2008. **4**(4): p. 256-63.
  95. Zhang, X., et al., *A potent small molecule inhibits polyglutamine aggregation in Huntington's disease neurons and suppresses neurodegeneration in vivo*. Proc Natl Acad Sci U S A, 2005. **102**(3): p. 892-7.
  96. Apostol, B.L., et al., *A cell-based assay for aggregation inhibitors as therapeutics of polyglutamine-repeat disease and validation in Drosophila*. Proc Natl Acad Sci U S A, 2003. **100**(10): p. 5950-5.
  97. Hotamisligil, G.S., *Inflammation and metabolic disorders*. Nature, 2006. **444**(7121): p. 860-7.
  98. Gomis Couto, A., et al., *Causes of unplanned hemodialysis initiation*. Nefrologia, 2011. **31**(6): p. 733-7.
  99. Dent, C.E. and G.R. Philpot, *Xanthinuria, an inborn error (or deviation) of metabolism*. Lancet, 1954. **266**(6804): p. 182-5.
  100. Ichida, K., et al., *Mutation of human molybdenum cofactor sulfurase gene is responsible for classical xanthinuria type II*. Biochem Biophys Res Commun, 2001. **282**(5): p. 1194-200.
  101. Watanabe, T., et al., *Deletion mutation in Drosophila ma-I homologous, putative molybdopterin cofactor sulfurase gene is associated with bovine xanthinuria type II*. J Biol Chem, 2000. **275**(29): p. 21789-92.
  102. Ramazzina, I., et al., *Completing the uric acid degradation pathway through phylogenetic comparison of whole genomes*. Nat Chem Biol, 2006. **2**(3): p. 144-8.
  103. Hadorn, E. and I. Schwinck, *[A mutant (rosy2) of Drosophila melanogaster without isoxanthopterin which is non-autonomous for the red eye pigments]*. Z Indukt Abstamm Vererbungsl, 1956. **87**(4): p. 528-53.
  104. Bonse, A., *[Studies on the chemical nature and formation of the urinary conglomerate in the Malpighian vessels of the rosy mutant of Drosophila melanogaster]*. Z Naturforsch B, 1967. **22**(10): p. 1027-9.
  105. Glassman, E. and H.K. Mitchell, *Mutants of Drosophila Melanogaster Deficient in Xanthine Dehydrogenase*. Genetics, 1959. **44**(2): p. 153-62.
  106. Kamleh, A., et al., *Metabolomic profiling using Orbitrap Fourier transform mass spectrometry with hydrophilic interaction chromatography: a method with wide applicability to analysis of biomolecules*. Rapid Commun Mass Spectrom, 2008. **22**(12): p. 1912-8.
  107. Kanbay, M., et al., *Uric acid in hypertension and renal disease: the chicken or the egg?* Blood Purif, 2010. **30**(4): p. 288-95.
  108. Feig, D.I., et al., *Uric acid and hypertension*. Curr Hypertens Rep, 2006. **8**(2): p. 111-5.

109. Khosla, U.M., et al., *Hyperuricemia induces endothelial dysfunction*. *Kidney Int*, 2005. **67**(5): p. 1739-42.
110. Gil-Campos, M., et al., *Uric acid is associated with features of insulin resistance syndrome in obese children at prepubertal stage*. *Nutr Hosp*, 2009. **24**(5): p. 607-13.
111. Pasalic, D., N. Marinkovic, and L. Feher-Turkovic, *Uric acid as one of the important factors in multifactorial disorders--facts and controversies*. *Biochem Med (Zagreb)*, 2012. **22**(1): p. 63-75.
112. Sun, X., et al., *Common variants related to serum uric acid concentrations are associated with glucose metabolism and insulin secretion in a Chinese population*. *PLoS One*, 2015. **10**(1): p. e0116714.
113. Kodama, S., et al., *Association between serum uric acid and development of type 2 diabetes*. *Diabetes Care*, 2009. **32**(9): p. 1737-42.
114. Jia, Z., et al., *Serum uric acid levels and incidence of impaired fasting glucose and type 2 diabetes mellitus: a meta-analysis of cohort studies*. *Diabetes Res Clin Pract*, 2013. **101**(1): p. 88-96.
115. Fan, H.Q., et al., *Association of serum uric acid with 2-hour postload glucose in Chinese with impaired fasting plasma glucose and/or HbA1c*. *PLoS One*, 2013. **8**(7): p. e67759.
116. Ohno, I., *Relationship between hyperuricemia and chronic kidney disease*. *Nucleosides Nucleotides Nucleic Acids*, 2011. **30**(12): p. 1039-44.
117. Pasalic, D., Marinkovic, N., Feher-Turkovic, L., *Uric acid as one of the important factors in multifactorial disorders – facts and controversies*. *Biochemia Medica*, 2012. **22**(1): p. 63-75.
118. Lee, Y., et al., *Mouse transthyretin-related protein is a hydrolase which degrades 5-hydroxyisourate, the end product of the uricase reaction*. *Mol Cells*, 2006. **22**(2): p. 141-5.
119. Tipton, P.A., *Urate to allantoin, specifically (S)-allantoin*. *Nat Chem Biol*, 2006. **2**(3): p. 124-5.
120. Becker, M.A., et al., *The urate-lowering efficacy and safety of febuxostat in the treatment of the hyperuricemia of gout: the CONFIRMS trial*. *Arthritis Res Ther*, 2010. **12**(2): p. R63.
121. Novak, O., et al., *Tissue specific profiling of the Arabidopsis thaliana auxin metabolome*. *Plant J*, 2012.
122. Nuringtyas, T.R., et al., *Differential tissue distribution of metabolites in Jacobaea vulgaris, Jacobaea aquatica and their crosses*. *Phytochemistry*, 2012. **78**: p. 89-97.
123. Bowen, B.P. and T.R. Northen, *Dealing with the unknown: metabolomics and metabolite atlases*. *J Am Soc Mass Spectrom*, 2010. **21**(9): p. 1471-6.
124. Adamski, J. and K. Suhre, *Metabolomics platforms for genome wide association studies--linking the genome to the metabolome*. *Curr Opin Biotechnol*, 2013. **24**(1): p. 39-47.
125. Putri, S.P., et al., *Current metabolomics: practical applications*. *J Biosci Bioeng*, 2013. **115**(6): p. 579-89.
126. Ryan, D. and K. Robards, *Metabolomics: The greatest omics of them all?* *Anal Chem*, 2006. **78**(23): p. 7954-8.
127. Xiao, J.F., B. Zhou, and H.W. Rensom, *Metabolite identification and quantitation in LC-MS/MS-based metabolomics*. *Trends Analyt Chem*, 2012. **32**: p. 1-14.
128. Nicholson, J.K., J.C. Lindon, and E. Holmes, *'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data*. *Xenobiotica*,

1999. **29**(11): p. 1181-9.
129. Weiss, R.H. and K. Kim, *Metabolomics in the study of kidney diseases*. Nat Rev Nephrol, 2012. **8**(1): p. 22-33.
130. Misra, B.B. and J.J. van der Hooft, *Updates in metabolomics tools and resources: 2014-2015*. Electrophoresis, 2016. **37**(1): p. 86-110.
131. Hyotylainen, T., *Novel methodologies in metabolic profiling with a focus on molecular diagnostic applications*. Expert Rev Mol Diagn, 2012. **12**(5): p. 527-38.
132. Yin, P. and G. Xu, *Current state-of-the-art of nontargeted metabolomics based on liquid chromatography-mass spectrometry with special emphasis in clinical applications*. J Chromatogr A, 2014. **1374**: p. 1-13.
133. Koal, T. and H.P. Deigner, *Challenges in mass spectrometry based targeted metabolomics*. Curr Mol Med, 2010. **10**(2): p. 216-26.
134. Mootha, V.K. and J.N. Hirschhorn, *Inborn variation in metabolism*. Nat Genet, 2010. **42**(2): p. 97-8.
135. Gika, H.G., I.D. Wilson, and G.A. Theodoridis, *LC-MS-based holistic metabolic profiling. Problems, limitations, advantages, and future perspectives*. J Chromatogr B Analyt Technol Biomed Life Sci, 2014. **966**: p. 1-6.
136. Lu, X., et al., *LC-MS-based metabolomics analysis*. J Chromatogr B Analyt Technol Biomed Life Sci, 2008. **866**(1-2): p. 64-76.
137. Rischer, H. and K.M. Oksman-Caldentey, *Unintended effects in genetically modified crops: revealed by metabolomics?* Trends Biotechnol, 2006. **24**(3): p. 102-4.
138. Schmid, M., et al., *A gene expression map of Arabidopsis thaliana development*. Nat Genet, 2005. **37**(5): p. 501-6.
139. E. S. Rosenblum , M.R.V., B. M. Braid, J. D. Moore, C. S. Friedman, R. S. Tjeerdema, *Characterizing the metabolic actions of natural stresses in the California red abalone, Haliotis rufescens using 1H NMR metabolomics*. Metabolomics, 2005. **1**(2): p. 199-209.
140. Saito, K. and F. Matsuda, *Metabolomics for functional genomics, systems biology, and biotechnology*. Annu Rev Plant Biol, 2010. **61**: p. 463-89.
141. Achcar, F., et al., *The silicon trypanosome: a test case of iterative model extension in systems biology*. Adv Microb Physiol, 2014. **64**: p. 115-43.
142. Barrett, M.P., B.M. Bakker, and R. Breitling, *Metabolomic systems biology of trypanosomes*. Parasitology, 2010. **137**(9): p. 1285-90.
143. Feala, J.D., et al., *Integrating metabolomics and phenomics with systems models of cardiac hypoxia*. Prog Biophys Mol Biol, 2008. **96**(1-3): p. 209-25.
144. Rochfort, S., *Metabolomics reviewed: a new "omics" platform technology for systems biology and implications for natural products research*. J Nat Prod, 2005. **68**(12): p. 1813-20.
145. Rhee, E.P. and R.E. Gerszten, *Metabolomics and cardiovascular biomarker discovery*. Clin Chem, 2012. **58**(1): p. 139-47.
146. Wang, Z., et al., *Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease*. Nature, 2011. **472**(7341): p. 57-63.
147. Connor, S.C., et al., *Integration of metabolomics and transcriptomics data to aid biomarker discovery in type 2 diabetes*. Mol Biosyst, 2010. **6**(5): p. 909-21.
148. Suhre, K., et al., *Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting*. PLoS One, 2010. **5**(11): p. e13953.
149. Teul, J., et al., *Improving metabolite knowledge in stable atherosclerosis patients by association and correlation of GC-MS and 1H NMR fingerprints*. J Proteome Res, 2009. **8**(12): p. 5580-9.



150. Sabatine, M.S., et al., *Metabolomic identification of novel biomarkers of myocardial ischemia*. *Circulation*, 2005. **112**(25): p. 3868-75.
151. Griffiths, W.J., et al., *Targeted metabolomics for biomarker discovery*. *Angew Chem Int Ed Engl*, 2010. **49**(32): p. 5426-45.
152. Ahmed, S.S., et al., *Metabolic profiling of Parkinson's disease: evidence of biomarker from gene expression analysis and rapid neural network detection*. *J Biomed Sci*, 2009. **16**: p. 63.
153. Underwood, B.R., et al., *Huntington disease patients and transgenic mice have similar pro-catabolic serum metabolite profiles*. *Brain*, 2006. **129**(Pt 4): p. 877-86.
154. Fiehn, O., et al., *Establishing reporting standards for metabolomic and metabonomic studies: a call for participation*. *OMICS*, 2006. **10**(2): p. 158-63.
155. Denkert, C., et al., *Mass spectrometry-based metabolic profiling reveals different metabolite patterns in invasive ovarian carcinomas and ovarian borderline tumors*. *Cancer Res*, 2006. **66**(22): p. 10795-804.
156. Fan, T.W., et al., *Altered regulation of metabolic pathways in human lung cancer discerned by (13)C stable isotope-resolved metabolomics (SIRM)*. *Mol Cancer*, 2009. **8**: p. 41.
157. Yoshida, M., et al., *[The challenge of disease diagnosis by metabolomics]*. *Fukuoka Igaku Zasshi*, 2010. **101**(11): p. 231-7.
158. Berg, M., et al., *LC-MS metabolomics from study design to data-analysis - using a versatile pathogen as a test case*. *Comput Struct Biotechnol J*, 2013. **4**: p. e201301002.
159. Creek, D.J., et al., *Metabolomic analysis of trypanosomatid protozoa*. *Mol Biochem Parasitol*, 2012. **181**(2): p. 73-84.
160. Yuan, J., et al., *Metabolomics-driven quantitative analysis of ammonia assimilation in E. coli*. *Mol Syst Biol*, 2009. **5**: p. 302.
161. Yoshida, S., et al., *Development of bottom-fermenting saccharomyces strains that produce high SO<sub>2</sub> levels, using integrated metabolome and transcriptome analysis*. *Appl Environ Microbiol*, 2008. **74**(9): p. 2787-96.
162. Raamsdonk, L.M., et al., *A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations*. *Nat Biotechnol*, 2001. **19**(1): p. 45-50.
163. Clasquin, M.F., et al., *Riboneogenesis in yeast*. *Cell*, 2011. **145**(6): p. 969-80.
164. Tweeddale, H., L. Notley-McRobb, and T. Ferenci, *Effect of slow growth on metabolism of Escherichia coli, as revealed by global metabolite pool ("metabolome") analysis*. *J Bacteriol*, 1998. **180**(19): p. 5109-16.
165. Shepherd, L.V., P. Fraser, and D. Stewart, *Metabolomics: a second-generation platform for crop and food analysis*. *Bioanalysis*, 2011. **3**(10): p. 1143-59.
166. Acharjee, A., et al., *Data integration and network reconstruction with ~omics data using Random Forest regression in potato*. *Anal Chim Acta*, 2011. **705**(1-2): p. 56-63.
167. Kende, A., et al., *Target list building for volatile metabolite profiling of fruit*. *J Chromatogr A*, 2010. **1217**(43): p. 6718-23.
168. Vaclavik, L., et al., *The use of high performance liquid chromatography-quadrupole time-of-flight mass spectrometry coupled to advanced data mining and chemometric tools for discrimination and classification of red wines according to their variety*. *Anal Chim Acta*, 2011. **685**(1): p. 45-51.
169. Rodrigues, J.A., et al., *Probing beer aging chemistry by nuclear magnetic resonance and multivariate analysis*. *Anal Chim Acta*, 2011. **702**(2): p. 178-87.
170. Hayashi, S., et al., *A novel application of metabolomics in vertebrate development*.

- Biochem Biophys Res Commun, 2009. **386**(1): p. 268-72.
171. Hayashi, S., et al., *Single-embryo metabolomics and systematic prediction of developmental stage in zebrafish*. Z Naturforsch C, 2011. **66**(3-4): p. 191-8.
  172. Ong, E.S., et al., *A multi-analytical approach for metabolomic profiling of zebrafish (*Danio rerio*) livers*. Mol Biosyst, 2009. **5**(3): p. 288-98.
  173. Pedersen, K.S., et al., *Metabolomic signatures of inbreeding at benign and stressful temperatures in *Drosophila melanogaster**. Genetics, 2008. **180**(2): p. 1233-43.
  174. Malmendal, A., et al., *Metabolomic profiling of heat stress: hardening and recovery of homeostasis in *Drosophila**. Am J Physiol Regul Integr Comp Physiol, 2006. **291**(1): p. R205-12.
  175. Overgaard, J., et al., *Metabolomic profiling of rapid cold hardening and cold shock in *Drosophila melanogaster**. J Insect Physiol, 2007. **53**(12): p. 1218-32.
  176. Coquin, L., et al., *Metabolomic and flux-balance analysis of age-related decline of hypoxia tolerance in *Drosophila* muscle tissue*. Mol Syst Biol, 2008. **4**: p. 233.
  177. Kamleh, M.A., et al., *Towards a platform for the metabonomic profiling of different strains of *Drosophila melanogaster* using liquid chromatography-Fourier transform mass spectrometry*. FEBS J, 2009. **276**(22): p. 6798-809.
  178. Hughes, S.L., et al., *The metabolomic responses of *Caenorhabditis elegans* to cadmium are largely independent of metallothionein status, but dominated by changes in cystathionine and phytochelatins*. J Proteome Res, 2009. **8**(7): p. 3512-9.
  179. Falk, M.J., et al., *Metabolic pathway profiling of mitochondrial respiratory chain mutants in *C. elegans**. Mol Genet Metab, 2008. **93**(4): p. 388-97.
  180. Nicholson, J.K., et al., *Metabonomics: a platform for studying drug toxicity and gene function*. Nat Rev Drug Discov, 2002. **1**(2): p. 153-61.
  181. Brindle, J.T., et al., *Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using <sup>1</sup>H-NMR-based metabonomics*. Nat Med, 2002. **8**(12): p. 1439-44.
  182. Leenders, J., M. Frederich, and P. de Tullio, *Nuclear magnetic resonance: a key metabolomics platform in the drug discovery process*. Drug Discov Today Technol, 2015. **13**: p. 39-46.
  183. Dettmer, K., P.A. Aronov, and B.D. Hammock, *Mass spectrometry-based metabolomics*. Mass Spectrom Rev, 2007. **26**(1): p. 51-78.
  184. Plumb, R.S., et al., *Metabonomics: the use of electrospray mass spectrometry coupled to reversed-phase liquid chromatography shows potential for the screening of rat urine in drug development*. Rapid Commun Mass Spectrom, 2002. **16**(20): p. 1991-6.
  185. Wilson, I.D., et al., *High resolution "ultra performance" liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies*. J Proteome Res, 2005. **4**(2): p. 591-8.
  186. Chintapalli, V.R., et al., *Mapping an atlas of tissue-specific *Drosophila melanogaster* metabolomes by high resolution mass spectrometry*. PLoS One, 2013. **8**(10): p. e78066.
  187. Kamleh, M.A., J.A. Dow, and D.G. Watson, *Applications of mass spectrometry in metabolomic studies of animal model and invertebrate systems*. Brief Funct Genomic Proteomic, 2009. **8**(1): p. 28-48.
  188. Scheltema, R.A., et al., *The potential of metabolomics for *Leishmania* research in the post-genomics era*. Parasitology, 2010. **137**(9): p. 1291-302.
  189. Theodoridis, G.A., et al., *Liquid chromatography-mass spectrometry based global*

- metabolite profiling: a review*. Anal Chim Acta, 2012. **711**: p. 7-16.
190. Malet-Martino, M. and U. Holzgrabe, *NMR techniques in biomedical and pharmaceutical analysis*. J Pharm Biomed Anal, 2011. **55**(1): p. 1-15.
  191. Wishart, D.S., et al., *HMDB: a knowledgebase for the human metabolome*. Nucleic Acids Res, 2009. **37**(Database issue): p. D603-10.
  192. Skogerson, K., et al., *The volatile compound BinBase mass spectral database*. BMC Bioinformatics, 2011. **12**: p. 321.
  193. Lindon, J.C., E. Holmes, and J.K. Nicholson, *So what's the deal with metabonomics?* Anal Chem, 2003. **75**(17): p. 384A-391A.
  194. Chou, Y.F., et al., *The effect of biomimetic apatite structure on osteoblast viability, proliferation, and gene expression*. Biomaterials, 2005. **26**(3): p. 285-95.
  195. Watson, D.G., *The potential of mass spectrometry for the global profiling of parasite metabolomes*. Parasitology, 2010. **137**(9): p. 1409-23.
  196. Glinski, M. and W. Weckwerth, *The role of mass spectrometry in plant systems biology*. Mass Spectrom Rev, 2006. **25**(2): p. 173-214.
  197. Perry, R.H., R.G. Cooks, and R.J. Noll, *Orbitrap mass spectrometry: instrumentation, ion motion and applications*. Mass Spectrom Rev, 2008. **27**(6): p. 661-99.
  198. Burgess, K., et al., *Semi-targeted analysis of metabolites using capillary-flow ion chromatography coupled to high-resolution mass spectrometry*. Rapid Commun Mass Spectrom, 2011. **25**(22): p. 3447-52.
  199. Koulman, A., et al., *High-resolution extracted ion chromatography, a new tool for metabolomics and lipidomics using a second-generation orbitrap mass spectrometer*. Rapid Commun Mass Spectrom, 2009. **23**(10): p. 1411-8.
  200. Gaspar, A., et al., *Trends in CE-MS 2005-2006*. Electrophoresis, 2008. **29**(1): p. 66-79.
  201. Honour, J.W., *Gas chromatography-mass spectrometry*. Methods Mol Biol, 2006. **324**: p. 53-74.
  202. Dunn, W.B., et al., *Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry*. Nat Protoc, 2011. **6**(7): p. 1060-83.
  203. Sreekumar, A., et al., *Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression*. Nature, 2009. **457**(7231): p. 910-4.
  204. Halket, J.M., et al., *Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS*. J Exp Bot, 2005. **56**(410): p. 219-43.
  205. Illig, T., et al., *A genome-wide perspective of genetic variation in human metabolism*. Nat Genet, 2010. **42**(2): p. 137-41.
  206. Zheng, L., et al., *Profiling of lipids in Leishmania donovani using hydrophilic interaction chromatography in combination with Fourier transform mass spectrometry*. Rapid Commun Mass Spectrom, 2010. **24**(14): p. 2074-82.
  207. Prakash, C., C.L. Shaffer, and A. Nedderman, *Analytical strategies for identifying drug metabolites*. Mass Spectrom Rev, 2007. **26**(3): p. 340-69.
  208. Giavalisco, P., et al., *<sup>13</sup>C isotope-labeled metabolomes allowing for improved compound annotation and relative quantification in liquid chromatography-mass spectrometry-based metabolomic research*. Anal Chem, 2009. **81**(15): p. 6546-51.
  209. Churchwell, M.I., et al., *Improving LC-MS sensitivity through increases in chromatographic performance: comparisons of UPLC-ES/MS/MS to HPLC-ES/MS/MS*. J Chromatogr B Analyt Technol Biomed Life Sci, 2005. **825**(2): p. 134-43.
  210. Bajad, S.U., et al., *Separation and quantitation of water soluble cellular*

- metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. J Chromatogr A, 2006. 1125(1): p. 76-88.*
211. Garcia, D.E., et al., *Separation and mass spectrometry in microbial metabolomics. Curr Opin Microbiol, 2008. 11(3): p. 233-9.*
  212. Liu, L., et al., *Fasting serum lipid and dehydroepiandrosterone sulfate as important metabolites for detecting isolated postchallenge diabetes: serum metabolomics via ultra-high-performance LC-MS. Clin Chem, 2013. 59(9): p. 1338-48.*
  213. Gaikwad, N.W., *Ultra performance liquid chromatography-tandem mass spectrometry method for profiling of steroid metabolome in human tissue. Anal Chem, 2013. 85(10): p. 4951-60.*
  214. Spagou, K., et al., *Hydrophilic interaction chromatography coupled to MS for metabonomic/metabolomic studies. J Sep Sci, 2010. 33(6-7): p. 716-27.*
  215. Gika, H.G., et al., *Quantitative profiling of polar primary metabolites using hydrophilic interaction ultrahigh performance liquid chromatography-tandem mass spectrometry. J Chromatogr A, 2012. 1259: p. 121-7.*
  216. Pesek, J.J., et al., *Aqueous normal-phase retention of nucleotides on silica hydride columns. J Chromatogr A, 2009. 1216(7): p. 1140-6.*
  217. t'Kindt, R., et al., *Towards an unbiased metabolic profiling of protozoan parasites: optimisation of a Leishmania sampling protocol for HILIC-orbitrap analysis. Anal Bioanal Chem, 2010. 398(5): p. 2059-69.*
  218. Wang, Y., X. Lu, and G. Xu, *Simultaneous separation of hydrophilic and hydrophobic compounds by using an online HILIC-RPLC system with two detectors. J Sep Sci, 2008. 31(9): p. 1564-72.*
  219. Fairchild, J.N., et al., *Two-dimensional liquid chromatography/mass spectrometry/mass spectrometry separation of water-soluble metabolites. J Chromatogr A, 2010. 1217(52): p. 8161-6.*
  220. Gika, H.G., et al., *Current practice of liquid chromatography-mass spectrometry in metabolomics and metabonomics. J Pharm Biomed Anal, 2014. 87: p. 12-25.*
  221. Yin, P., et al., *Preanalytical aspects and sample quality assessment in metabolomics studies of human blood. Clin Chem, 2013. 59(5): p. 833-45.*
  222. Yang, W., et al., *Liquid chromatography-tandem mass spectrometry-based plasma metabonomics delineate the effect of metabolites' stability on reliability of potential biomarkers. Anal Chem, 2013. 85(5): p. 2606-10.*
  223. Horai, H., et al., *MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom, 2010. 45(7): p. 703-14.*
  224. Smith, C.A., et al., *METLIN: a metabolite mass spectral database. Ther Drug Monit, 2005. 27(6): p. 747-51.*
  225. Codrea, M.C., et al., *Tools for computational processing of LC-MS datasets: a user's perspective. Comput Methods Programs Biomed, 2007. 86(3): p. 281-90.*
  226. Benton, H.P., et al., *XCMS2: processing tandem mass spectrometry data for metabolite identification and structural characterization. Anal Chem, 2008. 80(16): p. 6382-9.*
  227. Gowda, H., et al., *Interactive XCMS Online: simplifying advanced metabolomic data processing and subsequent statistical analyses. Anal Chem, 2014. 86(14): p. 6931-9.*
  228. Smith, C.A., et al., *XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem, 2006. 78(3): p. 779-87.*
  229. Katajamaa, M., J. Miettinen, and M. Oresic, *MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinformatics,*

2006. **22**(5): p. 634-6.
230. Pluskal, T., et al., *MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data*. BMC Bioinformatics, 2010. **11**: p. 395.
231. Bellew, M., et al., *A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS*. Bioinformatics, 2006. **22**(15): p. 1902-9.
232. Lommen, A., *MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing*. Anal Chem, 2009. **81**(8): p. 3079-86.
233. Tautenhahn, R., C. Bottcher, and S. Neumann, *Highly sensitive feature detection for high resolution LC/MS*. BMC Bioinformatics, 2008. **9**: p. 504.
234. Katajamaa, M. and M. Oresic, *Processing methods for differential analysis of LC/MS profile data*. BMC Bioinformatics, 2005. **6**: p. 179.
235. Coble, J.B. and C.G. Fraga, *Comparative evaluation of preprocessing freeware on chromatography/mass spectrometry data for signature discovery*. J Chromatogr A, 2014. **1358**: p. 155-64.
236. Yang, J., et al., *A data preprocessing strategy for metabolomics to reduce the mask effect in data analysis*. Front Mol Biosci, 2015. **2**: p. 4.
237. Nodzenski, M., et al., *Metabomxtr: an R package for mixture-model analysis of non-targeted metabolomics data*. Bioinformatics, 2014. **30**(22): p. 3287-8.
238. Creek, D.J., et al., *IDEOM: an Excel interface for analysis of LC-MS-based metabolomics data*. Bioinformatics, 2012. **28**(7): p. 1048-9.
239. Scheltema, R.A., et al., *PeakML/mzMatch: a file format, Java library, R library, and tool-chain for mass spectrometry data analysis*. Anal Chem, 2011. **83**(7): p. 2786-93.
240. Kuhl, C., et al., *CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets*. Anal Chem, 2012. **84**(1): p. 283-9.
241. Brown, M., et al., *Automated workflows for accurate mass-based putative metabolite identification in LC/MS-derived metabolomic datasets*. Bioinformatics, 2011. **27**(8): p. 1108-12.
242. Wishart, D.S., et al., *HMDB: the Human Metabolome Database*. Nucleic Acids Res, 2007. **35**(Database issue): p. D521-6.
243. Cui, Q., et al., *Metabolite identification via the Madison Metabolomics Consortium Database*. Nat Biotechnol, 2008. **26**(2): p. 162-4.
244. Kind, T. and O. Fiehn, *Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm*. BMC Bioinformatics, 2006. **7**: p. 234.
245. Vis, D.J., et al., *Network identification of hormonal regulation*. PLoS One, 2014. **9**(5): p. e96284.
246. De Livera, A.M., et al., *Statistical methods for handling unwanted variation in metabolomics data*. Anal Chem, 2015. **87**(7): p. 3606-15.
247. Taylor, J., et al., *Application of metabolomics to plant genotype discrimination using statistics and machine learning*. Bioinformatics, 2002. **18 Suppl 2**: p. S241-8.
248. Jansson, J., et al., *Metabolomics reveals metabolic biomarkers of Crohn's disease*. PLoS One, 2009. **4**(7): p. e6386.
249. Xi, B., et al., *Statistical analysis and modeling of mass spectrometry-based metabolomics data*. Methods Mol Biol, 2014. **1198**: p. 333-53.
250. Patterson, A.D., et al., *UPLC-ESI-TOFMS-based metabolomics and gene expression*

- dynamics inspector self-organizing metabolomic maps as tools for understanding the cellular response to ionizing radiation*. Anal Chem, 2008. **80**(3): p. 665-74.
251. Hamdalla, M.A., et al., *Metabolic pathway predictions for metabolomics: a molecular structure matching approach*. J Chem Inf Model, 2015. **55**(3): p. 709-18.
  252. Kaefer, A., et al., *MarVis-Pathway: integrative and exploratory pathway analysis of non-targeted metabolomics data*. Metabolomics, 2015. **11**(3): p. 764-777.
  253. Pon, A., et al., *Pathways with PathWhiz*. Nucleic Acids Res, 2015. **43**(W1): p. W552-9.
  254. Eichner, J., et al., *Integrated enrichment analysis and pathway-centered visualization of metabolomics, proteomics, transcriptomics, and genomics data by using the InCroMAP software*. J Chromatogr B Analyt Technol Biomed Life Sci, 2014. **966**: p. 77-82.
  255. Jankevics, A., et al., *Metabolomic analysis of a synthetic metabolic switch in Streptomyces coelicolor A3(2)*. Proteomics, 2011. **11**(24): p. 4622-31.
  256. Glaab, E. and R. Schneider, *RepExplore: addressing technical replicate variance in proteomics and metabolomics data analysis*. Bioinformatics, 2015. **31**(13): p. 2235-7.
  257. Chawade, A., E. Alexandersson, and F. Levander, *Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets*. J Proteome Res, 2014. **13**(6): p. 3114-20.
  258. Xia, J., et al., *MetaboAnalyst 3.0--making metabolomics more meaningful*. Nucleic Acids Res, 2015. **43**(W1): p. W251-7.
  259. Winkler, R., *MASSyPup--an 'out of the box' solution for the analysis of mass spectrometry data*. J Mass Spectrom, 2014. **49**(1): p. 37-42.
  260. Fukushima, A., et al., *Metabolomic Characterization of Knockout Mutants in Arabidopsis: Development of a Metabolite Profiling Database for Knockout Mutants in Arabidopsis*. Plant Physiol, 2014. **165**(3): p. 948-961.
  261. Alonso, A., S. Marsal, and A. Julia, *Analytical methods in untargeted metabolomics: state of the art in 2015*. Front Bioeng Biotechnol, 2015. **3**: p. 23.
  262. Sevin, D.C., et al., *Biological insights through nontargeted metabolomics*. Curr Opin Biotechnol, 2015. **34**: p. 1-8.
  263. Xu, Y., et al., *Evaluation of accurate mass and relative isotopic abundance measurements in the LTQ-orbitrap mass spectrometer for further metabolomics database building*. Anal Chem, 2010. **82**(13): p. 5490-501.
  264. Zelena, E., et al., *Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum*. Anal Chem, 2009. **81**(4): p. 1357-64.
  265. Yang, W.C., J. Adamec, and F.E. Regnier, *Enhancement of the LC/MS analysis of fatty acids through derivatization and stable isotope coding*. Anal Chem, 2007. **79**(14): p. 5150-7.
  266. Dai, W., et al., *Comprehensive and highly sensitive urinary steroid hormone profiling method based on stable isotope-labeling liquid chromatography-mass spectrometry*. Anal Chem, 2012. **84**(23): p. 10245-51.
  267. Yang, W.C., et al., *Enhancement of amino acid detection and quantification by electrospray ionization mass spectrometry*. Anal Chem, 2006. **78**(13): p. 4702-8.
  268. Ceglarek, U., et al., *Challenges and developments in tandem mass spectrometry based clinical metabolomics*. Mol Cell Endocrinol, 2009. **301**(1-2): p. 266-71.
  269. Tiller, P.R., et al., *High-throughput, accurate mass liquid chromatography/tandem mass spectrometry on a quadrupole time-of-flight system as a 'first-line' approach for metabolite identification studies*. Rapid Commun Mass Spectrom, 2008. **22**(7):

- p. 1053-61.
270. Hamdalla, M.A., R.A. Ammar, and S. Rajasekaran, *A molecular structure matching approach to efficient identification of endogenous mammalian biochemical structures*. BMC Bioinformatics, 2015. **16 Suppl 5**: p. S11.
  271. Allen, F., et al., *CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra*. Nucleic Acids Res, 2014. **42**(Web Server issue): p. W94-9.
  272. Zhou, J., et al., *HAMMER: automated operation of mass frontier to construct in silico mass spectral fragmentation libraries*. Bioinformatics, 2014. **30**(4): p. 581-3.
  273. Kanehisa, M., et al., *KEGG for integration and interpretation of large-scale molecular data sets*. Nucleic Acids Res, 2012. **40**(Database issue): p. D109-14.
  274. Haug, K., et al., *MetaboLights--an open-access general-purpose repository for metabolomics studies and associated meta-data*. Nucleic Acids Res, 2013. **41**(Database issue): p. D781-6.
  275. Wagele, B., et al., *MassTRIX reloaded: combined analysis and visualization of transcriptome and metabolome data*. PLoS One, 2012. **7**(7): p. e39860.
  276. Gika, H.G., et al., *Does the mass spectrometer define the marker? A comparison of global metabolite profiling data generated simultaneously via UPLC-MS on two different mass spectrometers*. Anal Chem, 2010. **82**(19): p. 8226-34.
  277. Lu, W., B.D. Bennett, and J.D. Rabinowitz, *Analytical strategies for LC-MS-based targeted metabolomics*. J Chromatogr B Analyt Technol Biomed Life Sci, 2008. **871**(2): p. 236-42.
  278. Lu, W., Y.K. Kwon, and J.D. Rabinowitz, *Isotope ratio-based profiling of microbial folates*. J Am Soc Mass Spectrom, 2007. **18**(5): p. 898-909.
  279. Quanbeck, S.M., et al., *Metabolomics as a Hypothesis-Generating Functional Genomics Tool for the Annotation of Arabidopsis thaliana Genes of "Unknown Function"*. Front Plant Sci, 2012. **3**: p. 15.
  280. Turi, C.E. and S.J. Murch, *Targeted and untargeted phytochemistry of Ligusticum canbyi: indoleamines, phthalides, antioxidant potential, and use of metabolomics as a hypothesis-generating technique for compound discovery*. Planta Med, 2013. **79**(14): p. 1370-9.
  281. Albeck, J.G., et al., *Collecting and organizing systematic sets of protein data*. Nat Rev Mol Cell Biol, 2006. **7**(11): p. 803-12.
  282. Breitling, R., *What is systems biology?* Front Physiol, 2010. **1**: p. 9.
  283. Hood, L., et al., *Systems biology and new technologies enable predictive and preventative medicine*. Science, 2004. **306**(5696): p. 640-3.
  284. Mustacchi, R., S. Hohmann, and J. Nielsen, *Yeast systems biology to unravel the network of life*. Yeast, 2006. **23**(3): p. 227-38.
  285. Stuart, L.M., et al., *A systems biology analysis of the Drosophila phagosome*. Nature, 2007. **445**(7123): p. 95-101.
  286. Miklos, G.L. and G.M. Rubin, *The role of the genome project in determining gene function: insights from model organisms*. Cell, 1996. **86**(4): p. 521-9.
  287. Ge, H., A.J. Walhout, and M. Vidal, *Integrating 'omic' information: a bridge between genomics and systems biology*. Trends Genet, 2003. **19**(10): p. 551-60.
  288. Bino, R.J., et al., *Potential of metabolomics as a functional genomics tool*. Trends Plant Sci, 2004. **9**(9): p. 418-25.
  289. Stitt, M. and A.R. Fernie, *From measurements of metabolites to metabolomics: an 'on the fly' perspective illustrated by recent studies of carbon-nitrogen interactions*. Curr Opin Biotechnol, 2003. **14**(2): p. 136-44.
  290. Oliver, S.G., *From DNA sequence to biological function*. Nature, 1996. **379**(6566):

- p. 597-600.
291. Joyce, A.R. and B.O. Palsson, *The model organism as a system: integrating 'omics' data sets*. Nat Rev Mol Cell Biol, 2006. **7**(3): p. 198-210.
  292. Fiehn, O., et al., *Metabolite profiling for plant functional genomics*. Nat Biotechnol, 2000. **18**(11): p. 1157-61.
  293. Weckwerth, W., et al., *Differential metabolic networks unravel the effects of silent plant phenotypes*. Proc Natl Acad Sci U S A, 2004. **101**(20): p. 7809-14.
  294. Goodacre, R., *Metabolomics shows the way to new discoveries*. Genome Biol, 2005. **6**(11): p. 354.
  295. Nielsen, J. and S. Oliver, *The next wave in metabolome analysis*. Trends Biotechnol, 2005. **23**(11): p. 544-6.
  296. Trethewey, R.N., *Gene discovery via metabolic profiling*. Curr Opin Biotechnol, 2001. **12**(2): p. 135-8.
  297. Fiehn, O. and W. Weckwerth, *Deciphering metabolic networks*. Eur J Biochem, 2003. **270**(4): p. 579-88.
  298. Kell, D.B., *Metabolomics and systems biology: making sense of the soup*. Curr Opin Microbiol, 2004. **7**(3): p. 296-307.
  299. Nicholson, J.K. and I.D. Wilson, *Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism*. Nat Rev Drug Discov, 2003. **2**(8): p. 668-76.
  300. Snoep, J.L., et al., *Towards building the silicon cell: a modular approach*. Biosystems, 2006. **83**(2-3): p. 207-16.
  301. Ideker, T., T. Galitski, and L. Hood, *A new approach to decoding life: systems biology*. Annu Rev Genomics Hum Genet, 2001. **2**: p. 343-72.
  302. Weston, A.D. and L. Hood, *Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine*. J Proteome Res, 2004. **3**(2): p. 179-96.
  303. Alcaraz, N., et al., *KeyPathwayMiner 4.0: condition-specific pathway analysis by combining multiple omics studies and networks with Cytoscape*. BMC Syst Biol, 2014. **8**: p. 99.
  304. King, Z.A., et al., *BiGG Models: A platform for integrating, standardizing and sharing genome-scale models*. Nucleic Acids Res, 2016. **44**(D1): p. D515-22.
  305. Bazzani, S., *Promise and reality in the expanding field of network interaction analysis: metabolic networks*. Bioinform Biol Insights, 2014. **8**: p. 83-91.
  306. Feist, A.M., et al., *Reconstruction of biochemical networks in microorganisms*. Nat Rev Microbiol, 2009. **7**(2): p. 129-43.
  307. Reed, J.L., et al., *Towards multidimensional genome annotation*. Nat Rev Genet, 2006. **7**(2): p. 130-41.
  308. Oberhardt, M.A., B.O. Palsson, and J.A. Papin, *Applications of genome-scale metabolic reconstructions*. Mol Syst Biol, 2009. **5**: p. 320.
  309. Bordbar, A., et al., *Constraint-based models predict metabolic and associated cellular functions*. Nat Rev Genet, 2014. **15**(2): p. 107-20.
  310. Duarte, N.C., et al., *Global reconstruction of the human metabolic network based on genomic and bibliomic data*. Proc Natl Acad Sci U S A, 2007. **104**(6): p. 1777-82.
  311. Rolfsson, O., B.O. Palsson, and I. Thiele, *The human metabolic reconstruction Recon 1 directs hypotheses of novel human metabolic functions*. BMC Syst Biol, 2011. **5**: p. 155.
  312. Folger, O., et al., *Predicting selective drug targets in cancer through metabolic networks*. Mol Syst Biol, 2011. **7**: p. 501.
  313. Frezza, C., et al., *Haem oxygenase is synthetically lethal with the tumour*



- suppressor fumarate hydratase*. *Nature*, 2011. **477**(7363): p. 225-8.
314. Plata, G., et al., *Reconstruction and flux-balance analysis of the Plasmodium falciparum metabolic network*. *Mol Syst Biol*, 2010. **6**: p. 408.
  315. Bonde, B.K., et al., *Differential producibility analysis (DPA) of transcriptomic data with metabolic networks: deconstructing the metabolic response of M. tuberculosis*. *PLoS Comput Biol*, 2011. **7**(6): p. e1002060.
  316. Metris, A., et al., *In vivo and in silico determination of essential genes of Campylobacter jejuni*. *BMC Genomics*, 2011. **12**: p. 535.
  317. Huthmacher, C., et al., *Antimalarial drug targets in Plasmodium falciparum predicted by stage-specific metabolic network analysis*. *BMC Syst Biol*, 2010. **4**: p. 120.
  318. Oberhardt, M.A., et al., *Reconciliation of genome-scale metabolic reconstructions for comparative systems analysis*. *PLoS Comput Biol*, 2011. **7**(3): p. e1001116.
  319. Orth, J.D. and B. Palsson, *Gap-filling analysis of the iJO1366 Escherichia coli metabolic network reconstruction for discovery of metabolic functions*. *BMC Syst Biol*, 2012. **6**: p. 30.
  320. Oberhardt, M.A., et al., *Genome-scale metabolic network analysis of the opportunistic pathogen Pseudomonas aeruginosa PAO1*. *J Bacteriol*, 2008. **190**(8): p. 2790-803.
  321. Puchalka, J., et al., *Genome-scale reconstruction and analysis of the Pseudomonas putida KT2440 metabolic network facilitates applications in biotechnology*. *PLoS Comput Biol*, 2008. **4**(10): p. e1000210.
  322. Sigurdsson, M.I., et al., *A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1*. *BMC Syst Biol*, 2010. **4**: p. 140.
  323. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes*. *Nucleic Acids Res*, 2000. **28**(1): p. 27-30.
  324. Schomburg, I., et al., *BRENDA: a resource for enzyme data and metabolic information*. *Trends Biochem Sci*, 2002. **27**(1): p. 54-6.
  325. Caspi, R. and P.D. Karp, *Using the MetaCyc pathway database and the BioCyc database collection*. *Curr Protoc Bioinformatics*, 2007. **Chapter 1**: p. Unit1 17.
  326. Thiele, I. and B.O. Palsson, *A protocol for generating a high-quality genome-scale metabolic reconstruction*. *Nat Protoc*, 2010. **5**(1): p. 93-121.
  327. Cicek, A.E., K. Roeder, and G. Ozsoyoglu, *MIRA: mutual information-based reporter algorithm for metabolic networks*. *Bioinformatics*, 2014. **30**(12): p. i175-84.
  328. Grapov, D., K. Wanichthanarak, and O. Fiehn, *MetaMapR: pathway independent metabolomic network analysis incorporating unknowns*. *Bioinformatics*, 2015. **31**(16): p. 2757-60.
  329. Carazzolle, M.F., et al., *IIS--Integrated Interactome System: a web-based platform for the annotation, analysis and visualization of protein-metabolite-gene-drug interactions by integrating a variety of data sources and tools*. *PLoS One*, 2014. **9**(6): p. e100385.
  330. Becker, S.A., et al., *Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox*. *Nat Protoc*, 2007. **2**(3): p. 727-38.
  331. Schellenberger, J., et al., *Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0*. *Nat Protoc*, 2011. **6**(9): p. 1290-307.
  332. Kauffman, K.J., P. Prakash, and J.S. Edwards, *Advances in flux balance analysis*. *Curr Opin Biotechnol*, 2003. **14**(5): p. 491-6.
  333. Schuetz, R., L. Kuepfer, and U. Sauer, *Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli*. *Mol Syst Biol*, 2007. **3**: p. 119.

334. Varma, A. and B.O. Palsson, *Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110*. Appl Environ Microbiol, 1994. **60**(10): p. 3724-31.
335. Orth, J.D. and B.O. Palsson, *Systematizing the generation of missing metabolic knowledge*. Biotechnol Bioeng, 2010. **107**(3): p. 403-12.
336. Chen, L. and D. Vitkup, *Distribution of orphan metabolic activities*. Trends Biotechnol, 2007. **25**(8): p. 343-8.
337. Karp, P.D., *Call for an enzyme genomics initiative*. Genome Biol, 2004. **5**(8): p. 401.
338. Osterman, A. and R. Overbeek, *Missing genes in metabolic pathways: a comparative genomics approach*. Curr Opin Chem Biol, 2003. **7**(2): p. 238-51.
339. Hanson, A.D., et al., *'Unknown' proteins and 'orphan' enzymes: the missing half of the engineering parts list--and how to find it*. Biochem J, 2010. **425**(1): p. 1-11.
340. Reed, J.L., et al., *Systems approach to refining genome annotation*. Proc Natl Acad Sci U S A, 2006. **103**(46): p. 17480-4.
341. Satish Kumar, V., M.S. Dasika, and C.D. Maranas, *Optimization based automated curation of metabolic reconstructions*. BMC Bioinformatics, 2007. **8**: p. 212.
342. Kumar, V.S. and C.D. Maranas, *GrowMatch: an automated method for reconciling in silico/in vivo growth predictions*. PLoS Comput Biol, 2009. **5**(3): p. e1000308.
343. Paley, S.M. and P.D. Karp, *Evaluation of computational metabolic-pathway predictions for Helicobacter pylori*. Bioinformatics, 2002. **18**(5): p. 715-24.
344. Osterlund, T., et al., *Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling*. BMC Syst Biol, 2013. **7**: p. 36.
345. Vanholme, R., et al., *A systems biology view of responses to lignin biosynthesis perturbations in Arabidopsis*. Plant Cell, 2012. **24**(9): p. 3506-29.
346. Kotze, H.L., et al., *A novel untargeted metabolomics correlation-based network analysis incorporating human metabolic reconstructions*. BMC Syst Biol, 2013. **7**: p. 107.
347. Edison, A.S., et al., *The Time Is Right to Focus on Model Organism Metabolomes*. Metabolites, 2016. **6**(1).
348. Feala, J.D., et al., *Flexibility in energy metabolism supports hypoxia tolerance in Drosophila flight muscle: metabolomic and computational systems analysis*. Mol Syst Biol, 2007. **3**: p. 99.
349. Feala, J.D., et al., *Metabolism as means for hypoxia adaptation: metabolic profiling and flux balance analysis*. BMC Syst Biol, 2009. **3**: p. 91.
350. van Iterson, M., et al., *Relative power and sample size analysis on gene expression profiling data*. BMC Genomics, 2009. **10**: p. 439.
351. Ainsworth, C., S. Wan, and H. Skaer, *Coordinating cell fate and morphogenesis in Drosophila renal tubules*. Philos Trans R Soc Lond B Biol Sci, 2000. **355**(1399): p. 931-7.
352. Dow, J.A., et al., *The malpighian tubules of Drosophila melanogaster: a novel phenotype for studies of fluid secretion and its control*. J Exp Biol, 1994. **197**: p. 421-8.
353. Creek, D.J., et al., *Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: improved metabolite identification by retention time prediction*. Anal Chem, 2011. **83**(22): p. 8703-10.
354. Curto, R., et al., *Mathematical models of purine metabolism in man*. Math Biosci, 1998. **151**(1): p. 1-49.
355. Kahn, K. and P.A. Tipton, *Spectroscopic characterization of intermediates in the urate oxidase reaction*. Biochemistry, 1998. **37**(33): p. 11651-9.
356. Kahn, K. and P.A. Tipton, *Kinetic mechanism and cofactor content of soybean root*

- nodule urate oxidase*. Biochemistry, 1997. **36**(16): p. 4731-8.
357. Todd, C.D., et al., *Update on ureide degradation in legumes*. J Exp Bot, 2006. **57**(1): p. 5-12.
  358. Chen, B., et al., *Linearly Polymerized Benzene Arrays As Intermediates, Tracing Pathways to Carbon Nanotubes*. J Am Chem Soc, 2015.
  359. Sarma, A.D., et al., *Identification and purification of hydroxyisourate hydrolase, a novel ureide-metabolizing enzyme*. J Biol Chem, 1999. **274**(48): p. 33863-5.
  360. Caetano-Anolles, G., H.S. Kim, and J.E. Mittenthal, *The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture*. Proc Natl Acad Sci U S A, 2007. **104**(22): p. 9358-63.
  361. Oda, M., et al., *Loss of urate oxidase activity in hominoids and its evolutionary implications*. Mol Biol Evol, 2002. **19**(5): p. 640-53.
  362. Vogels, G.D. and C. Van der Drift, *Degradation of purines and pyrimidines by microorganisms*. Bacteriol Rev, 1976. **40**(2): p. 403-68.
  363. Lee, K.W. and A.H. Roush, *Allantoinase Assays and Their Application to Yeast and Soybean Allantoinases*. Arch Biochem Biophys, 1964. **108**: p. 460-7.
  364. Mulrooney, S.B. and R.P. Hausinger, *Metal ion dependence of recombinant Escherichia coli allantoinase*. J Bacteriol, 2003. **185**(1): p. 126-34.
  365. Santos, C.X., E.I. Anjos, and O. Augusto, *Uric acid oxidation by peroxynitrite: multiple reactions, free radical formation, and amplification of lipid oxidation*. Arch Biochem Biophys, 1999. **372**(2): p. 285-94.
  366. Bender, W., P. Spierer, and D.S. Hogness, *Chromosomal walking and jumping to isolate DNA from the Ace and rosy loci and the bithorax complex in Drosophila melanogaster*. J Mol Biol, 1983. **168**(1): p. 17-33.
  367. Kral, L.G., et al., *Cloning a cDNA for Drosophila melanogaster urate oxidase*. Gene, 1986. **45**(2): p. 131-7.
  368. Friedman, T.B. and C.R. Merrill, *A microradiochemical assay for urate oxidase*. Anal Biochem, 1973. **55**(1): p. 292-6.
  369. Wallrath, L.L., J.B. Burnett, and T.B. Friedman, *Molecular characterization of the Drosophila melanogaster urate oxidase gene, an ecdysone-repressible gene expressed only in the malpighian tubules*. Mol Cell Biol, 1990. **10**(10): p. 5114-27.
  370. Friedman, T.B., *Observations on the regulation of uricase activity during development of Drosophila melanogaster*. Biochem Genet, 1973. **8**(1): p. 37-45.
  371. Friedman, T.B. and D.H. Johnson, *Temporal control of urate oxidase activity in Drosophila: evidence of an autonomous timer in malpighian tubules*. Science, 1977. **197**(4302): p. 477-9.
  372. Raychaudhuri, A. and P.A. Tipton, *Cloning and expression of the gene for soybean hydroxyisourate hydrolase. Localization and implications for function and mechanism*. Plant Physiol, 2002. **130**(4): p. 2061-8.
  373. Gournas, C., et al., *Completing the purine utilisation pathway of Aspergillus nidulans*. Fungal Genet Biol, 2011. **48**(8): p. 840-8.
  374. Kasai, K., N. Nishiyama, and K. Yamauchi, *Characterization of Oncorhynchus mykiss 5-hydroxyisourate hydrolase/transthyretin superfamily: evolutionary and functional analyses*. Gene, 2013. **531**(2): p. 326-36.
  375. Zanotti, G., et al., *Structure of zebra fish HUase: insights into evolution of an enzyme to a hormone transporter*. J Mol Biol, 2006. **363**(1): p. 1-9.
  376. Hennebry, S.C., et al., *The crystal structure of the transthyretin-like protein from Salmonella dublin, a prokaryote 5-hydroxyisourate hydrolase*. J Mol Biol, 2006. **359**(5): p. 1389-99.
  377. Eneqvist, T., et al., *The transthyretin-related protein family*. Eur J Biochem, 2003.

- 270**(3): p. 518-32.
378. Lundberg, E., et al., *The transthyretin-related protein: structural investigation of a novel protein family*. J Struct Biol, 2006. **155**(3): p. 445-57.
  379. Lee, I.R., et al., *Characterization of the complete uric acid degradation pathway in the fungal pathogen Cryptococcus neoformans*. PLoS One, 2013. **8**(5): p. e64292.
  380. French, J.B. and S.E. Ealick, *Structural and kinetic insights into the mechanism of 5-hydroxyisourate hydrolase from Klebsiella pneumoniae*. Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 8): p. 671-7.
  381. Keebaugh, A.C. and J.W. Thomas, *The genomes of the South American opossum (Monodelphis domestica) and platypus (Ornithorhynchus anatinus) encode a more complete purine catabolic pathway than placental mammals*. Comp Biochem Physiol Part D Genomics Proteomics, 2009. **4**(3): p. 174-8.
  382. Matiollo, C., et al., *A transthyretin-related protein is functionally expressed in Herbaspirillum seropedicae*. Biochem Biophys Res Commun, 2009. **387**(4): p. 712-6.
  383. Li, Z., et al., *Identification and bioactivity analysis of transthyretin-like protein in amphioxus: a case demonstrating divergent evolution from an enzyme to a hormone distributor*. Comp Biochem Physiol B Biochem Mol Biol, 2013. **164**(3): p. 143-50.
  384. Lee, Y., et al., *Transthyretin-related proteins function to facilitate the hydrolysis of 5-hydroxyisourate, the end product of the uricase reaction*. FEBS Lett, 2005. **579**(21): p. 4769-74.
  385. Schreiber, G. and S.J. Richardson, *The evolution of gene expression, structure and function of transthyretin*. Comp Biochem Physiol B Biochem Mol Biol, 1997. **116**(2): p. 137-60.
  386. Prapunpoj, P., et al., *Evolution of structure, ontogeny of gene expression, and function of Xenopus laevis transthyretin*. Am J Physiol Regul Integr Comp Physiol, 2000. **279**(6): p. R2026-41.
  387. Richardson, S.J., *The evolution of transthyretin synthesis in vertebrate liver, in primitive eukaryotes and in bacteria*. Clin Chem Lab Med, 2002. **40**(12): p. 1191-9.
  388. Hennebry, S.C., et al., *Structural and functional evolution of transthyretin and transthyretin-like proteins*. Proteins, 2006. **64**(4): p. 1024-45.
  389. Raychaudhuri, A. and P.A. Tipton, *A familiar motif in a new context: the catalytic mechanism of hydroxyisourate hydrolase*. Biochemistry, 2003. **42**(22): p. 6848-52.
  390. Power, D.M., et al., *Evolution of the thyroid hormone-binding protein, transthyretin*. Gen Comp Endocrinol, 2000. **119**(3): p. 241-55.
  391. Hayashi, S., S. Fujiwara, and T. Noguchi, *Evolution of urate-degrading enzymes in animal peroxisomes*. Cell Biochem Biophys, 2000. **32 Spring**: p. 123-9.
  392. Nakai, K. and M. Kanehisa, *A knowledge base for predicting protein localization sites in eukaryotic cells*. Genomics, 1992. **14**(4): p. 897-911.
  393. Nielsen, H., et al., *Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites*. Protein Eng, 1997. **10**(1): p. 1-6.
  394. Subramani, S., A. Koller, and W.B. Snyder, *Import of peroxisomal matrix and membrane proteins*. Annu Rev Biochem, 2000. **69**: p. 399-418.
  395. Reynolds, P.H., et al., *Enzymes of amide and ureide biogenesis in developing soybean nodules*. Plant Physiol, 1982. **69**(6): p. 1334-8.
  396. Chintapalli, V.R., et al., *Functional correlates of positional and gender-specific renal asymmetry in Drosophila*. PLoS One, 2012. **7**(4): p. e32577.
  397. Dow, J.A. and M.F. Romero, *Drosophila provides rapid modeling of renal development, function, and disease*. Am J Physiol Renal Physiol, 2010. **299**(6): p.

F1237-44.

398. Sozen, M.A., et al., *Functional domains are specified to single-cell resolution in a Drosophila epithelium*. Proc Natl Acad Sci U S A, 1997. **94**(10): p. 5207-12.
399. Beyenbach, K.W., H. Skaer, and J.A. Dow, *The developmental, molecular, and transport biology of Malpighian tubules*. Annu Rev Entomol, 2010. **55**: p. 351-74.
400. H, S. and M.-A. A, *The wingless product is required for cell proliferation in the Malpighian tubule anlage of Drosophila melanogaster*. Development 1992. **116**: p. 745-754.
401. McCarty, D.J. and J.L. Hollander, *Identification of urate crystals in gouty synovial fluid*. Ann Intern Med, 1961. **54**: p. 452-60.
402. Keith, T.P., et al., *Sequence of the structural gene for xanthine dehydrogenase (rosy locus) in Drosophila melanogaster*. Genetics, 1987. **116**(1): p. 67-73.
403. Mitchell, H.K. and E. Glassman, *Hypoxanthine in rosy and maroon-like mutants of Drosophila melanogaster*. Science, 1959. **129**(3344): p. 268.
404. Hobani, Y.H., *Metabolomic analyses of Drosophila models for human renal disease*. PhD thesis, 2012
405. Davies, S.M., et al., *Idiopathic hyperammonemia: a frequently lethal complication of bone marrow transplantation*. Bone Marrow Transplant, 1996. **17**(6): p. 1119-25.
406. Gausz, J., et al., *Genetic Characterization of the 87c Region of the Third Chromosome of DROSOPHILA MELANOGASTER*. Genetics, 1979. **93**(4): p. 917-34.
407. Dow, J.T. and S.A. Davies, *Integrative physiology and functional genomics of epithelial function in a genetic model organism*. Physiol Rev, 2003. **83**(3): p. 687-729.
408. Karet, F.E., et al., *Mutations in the gene encoding B1 subunit of H<sup>+</sup>-ATPase cause renal tubular acidosis with sensorineural deafness*. Nat Genet, 1999. **21**(1): p. 84-90.
409. Jack, J. and G. Myette, *The genes raw and ribbon are required for proper shape of tubular epithelial tissues in Drosophila*. Genetics, 1997. **147**(1): p. 243-53.
410. Dougan, S. and S. DiNardo, *Drosophila wingless generates cell type diversity among engrailed expressing cells*. Nature, 1992. **360**(6402): p. 347-50.
411. Chintapalli, V.R., et al., *Data-mining the FlyAtlas online resource to identify core functional motifs across transporting epithelia*. BMC Genomics, 2013. **14**: p. 518.
412. Cochat, P., et al., *Nephrolithiasis related to inborn metabolic diseases*. Pediatr Nephrol, 2010. **25**(3): p. 415-24.
413. Watts, R.W., *Metabolic causes of renal stone formation*. Postgrad Med J, 1977. **53 Suppl 2**: p. 7-24.
414. Evans, J.M., et al., *Sulphonylurea sensitivity and enriched expression implicate inward rectifier K<sup>+</sup> channels in Drosophila melanogaster renal function*. J Exp Biol, 2005. **208**(Pt 19): p. 3771-83.
415. Dow, J.A. and S.A. Davies, *The Malpighian tubule: rapid insights from post-genomic biology*. J Insect Physiol, 2006. **52**(4): p. 365-78.
416. Davies, S.A., et al., *Analysis and inactivation of vha55, the gene encoding the vacuolar ATPase B-subunit in Drosophila melanogaster reveals a larval lethal phenotype*. J Biol Chem, 1996. **271**(48): p. 30677-84.
417. Hirata, T., et al., *Ion and solute transport by Prestin in Drosophila and Anopheles*. J Insect Physiol, 2012. **58**(4): p. 563-9.
418. Landry, G.M., et al., *Sulfate and thiosulfate inhibit oxalate transport via a dPrestin (Slc26a6)-dependent mechanism in an insect model of calcium oxalate nephrolithiasis*. Am J Physiol Renal Physiol, 2016. **310**(2): p. F152-9.

419. Becker, B.F., *Towards the physiological function of uric acid*. Free Radic Biol Med, 1993. **14**(6): p. 615-31.
420. Cutler, R.G., *Urate and ascorbate: their possible roles as antioxidants in determining longevity of mammalian species*. Arch Gerontol Geriatr, 1984. **3**(4): p. 321-48.
421. Ames, B.N., et al., *Uric acid provides an antioxidant defense in humans against oxidant- and radical-caused aging and cancer: a hypothesis*. Proc Natl Acad Sci U S A, 1981. **78**(11): p. 6858-62.
422. Bergamini, C., et al., *Oxidative stress and hyperuricaemia: pathophysiology, clinical relevance, and therapeutic implications in chronic heart failure*. Eur J Heart Fail, 2009. **11**(5): p. 444-52.
423. Gersch, C., et al., *Reactions of peroxynitrite with uric acid: formation of reactive intermediates, alkylated products and triuret, and in vivo production of triuret under conditions of oxidative stress*. Nucleosides Nucleotides Nucleic Acids, 2009. **28**(2): p. 118-49.
424. Glantzounis, G.K., et al., *Uric acid and oxidative stress*. Curr Pharm Des, 2005. **11**(32): p. 4145-51.
425. Jia, L., et al., *Hyperuricemia causes pancreatic beta-cell death and dysfunction through NF-kappaB signaling pathway*. PLoS One, 2013. **8**(10): p. e78284.
426. Modan, M., et al., *Elevated serum uric acid--a facet of hyperinsulinaemia*. Diabetologia, 1987. **30**(9): p. 713-8.
427. Zhang, Y., et al., *Uric acid induces oxidative stress and growth inhibition by activating adenosine monophosphate-activated protein kinase and extracellular signal-regulated kinase signal pathways in pancreatic beta cells*. Mol Cell Endocrinol, 2013. **375**(1-2): p. 89-96.
428. Yu, B.P., *Aging and oxidative stress: modulation by dietary restriction*. Free Radic Biol Med, 1996. **21**(5): p. 651-68.
429. Harman, D., *Aging: a theory based on free radical and radiation chemistry*. J Gerontol, 1956. **11**(3): p. 298-300.
430. Kim, Y.S., et al., *Role of xanthine dehydrogenase and aging on the innate immune response of Drosophila*. J Am Aging Assoc, 2001. **24**(4): p. 187-93.
431. Waring, W.S., *Uric acid: an important antioxidant in acute ischaemic stroke*. QJM, 2002. **95**(10): p. 691-3.
432. Moody, C.S. and H.M. Hassan, *Mutagenicity of oxygen free radicals*. Proc Natl Acad Sci U S A, 1982. **79**(9): p. 2855-9.
433. Sohal, R.S., et al., *Effect of age and ambient temperature on n-pentane production in adult housefly, Musca domestica*. Mech Ageing Dev, 1985. **29**(3): p. 317-26.
434. Sagai, M. and T. Ichinose, *Age-related changes in lipid peroxidation as measured by ethane, ethylene, butane and pentane in respired gases of rats*. Life Sci, 1980. **27**(9): p. 731-8.
435. Nieto, F.J., et al., *Uric acid and serum antioxidant capacity: a reaction to atherosclerosis?* Atherosclerosis, 2000. **148**(1): p. 131-9.
436. Gardner, I.D., *The effect of aging on susceptibility to infection*. Rev Infect Dis, 1980. **2**(5): p. 801-10.
437. Humphreys, J.M., A.J. Hilliker, and J.P. Phillips, *Paraquat selection identifies X-linked oxygen defense genes in Drosophila melanogaster*. Genome, 1993. **36**(1): p. 162-5.
438. Le Bourg, E., *Oxidative stress, aging and longevity in Drosophila melanogaster*. FEBS Lett, 2001. **498**(2-3): p. 183-6.
439. Sohal, R.S., L. Arnold, and W.C. Orr, *Effect of age on superoxide dismutase,*

- catalase, glutathione reductase, inorganic peroxides, TBA-reactive material, GSH/GSSG, NADPH/NADP+ and NADH/NAD+ in Drosophila melanogaster*. Mech Ageing Dev, 1990. **56**(3): p. 223-35.
440. Schulz, E., T. Gori, and T. Munzel, *Oxidative stress and endothelial dysfunction in hypertension*. Hypertens Res, 2011. **34**(6): p. 665-73.
  441. Higgins, P., L.D. Ferguson, and M.R. Walters, *Xanthine oxidase inhibition for the treatment of stroke disease: a novel therapeutic approach*. Expert Rev Cardiovasc Ther, 2011. **9**(4): p. 399-401.
  442. Farrington, J.A., et al., *Bipyridylium quaternary salts and related compounds. V. Pulse radiolysis studies of the reaction of paraquat radical with oxygen. Implications for the mode of action of bipyridyl herbicides*. Biochim Biophys Acta, 1973. **314**(3): p. 372-81.
  443. Tolmasoff, J.M., T. Ono, and R.G. Cutler, *Superoxide dismutase: correlation with life-span and specific metabolic rate in primate species*. Proc Natl Acad Sci U S A, 1980. **77**(5): p. 2777-81.
  444. Tatar, M., S. Post, and K. Yu, *Nutrient control of Drosophila longevity*. Trends Endocrinol Metab, 2014. **25**(10): p. 509-17.
  445. An, P.N., et al., *Metabolome analysis of Drosophila melanogaster during embryogenesis*. PLoS One, 2014. **9**(8): p. e99519.
  446. Bratty, M.A., et al., *Metabolomic profiling reveals that Drosophila melanogaster larvae with the y mutation have altered lysine metabolism*. FEBS Open Bio, 2012. **2**: p. 217-21.
  447. Tennessen, J.M., et al., *Coordinated metabolic transitions during Drosophila embryogenesis and the onset of aerobic glycolysis*. G3 (Bethesda), 2014. **4**(5): p. 839-50.
  448. Heinrichsen, E.T., et al., *Metabolic and transcriptional response to a high-fat diet in Drosophila melanogaster*. Mol Metab, 2014. **3**(1): p. 42-54.
  449. Laye, M.J., et al., *The effects of age and dietary restriction on the tissue-specific metabolome of Drosophila*. Aging Cell, 2015. **14**(5): p. 797-808.
  450. Williams, S., et al., *Metabolomic and Gene Expression Profiles Exhibit Modular Genetic and Dietary Structure Linking Metabolic Syndrome Phenotypes in Drosophila*. G3 (Bethesda), 2015. **5**(12): p. 2817-29.
  451. Luan, H., et al., *LC-MS-based urinary metabolite signatures in idiopathic Parkinson's disease*. J Proteome Res, 2015. **14**(1): p. 467-78.
  452. Shukla, A.K., et al., *Metabolomic Analysis Provides Insights on Paraquat-Induced Parkinson-Like Symptoms in Drosophila melanogaster*. Mol Neurobiol, 2016. **53**(1): p. 254-69.
  453. Gogna, N., et al., *NMR-based investigation of the Drosophila melanogaster metabolome under the influence of daily cycles of light and temperature*. Mol Biosyst, 2015. **11**(12): p. 3305-15.
  454. Hariharan, R., et al., *Invariance and plasticity in the Drosophila melanogaster metabolomic network in response to temperature*. BMC Syst Biol, 2014. **8**: p. 139.
  455. Williams, C.M., et al., *Cold adaptation shapes the robustness of metabolic networks in Drosophila melanogaster*. Evolution, 2014. **68**(12): p. 3505-23.
  456. Hoffman, J.M., et al., *Effects of age, sex, and genotype on high-sensitivity metabolomic profiles in the fruit fly, Drosophila melanogaster*. Aging Cell, 2014. **13**(4): p. 596-604.
  457. Carvalho, M., et al., *Effects of diet and development on the Drosophila lipidome*. Mol Syst Biol, 2012. **8**: p. 600.
  458. Xia, J., et al., *MetaboAnalyst: a web server for metabolomic data analysis and*

- interpretation*. Nucleic Acids Res, 2009. **37**(Web Server issue): p. W652-60.
459. Tearle, R., *Tissue specific effects of ommochrome pathway mutations in Drosophila melanogaster*. Genet Res, 1991. **57**(3): p. 257-66.
  460. Sullivan, D.T., et al., *Genetic and functional analysis of tryptophan transport in Malpighian tubules of Drosophila*. Biochem Genet, 1980. **18**(11-12): p. 1109-30.
  461. Kim, K., J. Park, and S. Rhee, *Structural and functional basis for (S)-allantoin formation in the ureide pathway*. J Biol Chem, 2007. **282**(32): p. 23457-64.
  462. Pessoa, J., et al., *Functional characterization of Arabidopsis thaliana transthyretin-like protein*. BMC Plant Biol, 2010. **10**: p. 30.
  463. Jung, D.K., et al., *Structural and functional analysis of PucM, a hydrolase in the ureide pathway and a member of the transthyretin-related protein family*. Proc Natl Acad Sci U S A, 2006. **103**(26): p. 9790-5.
  464. Lamberto, I., et al., *Conserved alternative splicing of Arabidopsis transthyretin-like determines protein localization and S-allantoin synthesis in peroxisomes*. Plant Cell, 2010. **22**(5): p. 1564-74.
  465. Wikoff, W.R., et al., *Metabolomic analysis of the cerebrospinal fluid reveals changes in phospholipase expression in the CNS of SIV-infected macaques*. J Clin Invest, 2008. **118**(7): p. 2661-9.
  466. Ezure, T., et al., *Cell-free protein synthesis system prepared from insect cells by freeze-thawing*. Biotechnol Prog, 2006. **22**(6): p. 1570-7.
  467. Rosenberg, A.H. and F.W. Studier, *T7 RNA polymerase can direct expression of influenza virus cap-binding protein (PB2) in Escherichia coli*. Gene, 1987. **59**(2-3): p. 191-200.
  468. Studier, F.W., et al., *Use of T7 RNA polymerase to direct expression of cloned genes*. Methods Enzymol, 1990. **185**: p. 60-89.
  469. Dubendorff, J.W. and F.W. Studier, *Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor*. J Mol Biol, 1991. **219**(1): p. 45-59.
  470. Schneider, I., *Cell lines derived from late embryonic stages of Drosophila melanogaster*. J Embryol Exp Morphol, 1972. **27**(2): p. 353-65.
  471. Davies, S.A., et al., *Cell signalling mechanisms for insect stress tolerance*. J Exp Biol, 2014. **217**(Pt 1): p. 119-28.
  472. Ideker, T., et al., *Integrated genomic and proteomic analyses of a systematically perturbed metabolic network*. Science, 2001. **292**(5518): p. 929-34.
  473. Saez, L. and M.W. Young, *Regulation of nuclear entry of the Drosophila clock proteins period and timeless*. Neuron, 1996. **17**(5): p. 911-20.



# Appendices

## Appendix I: Fly food recipe

Per 1 litre of water
10 g agar
15 g sucrose
30 g glucose
35 g dried yeast
15 g maize meal
10 g wheat germ
30 g treacle
10 g soya flour
Bring to boil, simmer for 10 min, cool to 70°C, leave for 20 min, and add the following:
10 ml Nipagin (25 g Nipagin M (Tegosept M, p-hydroxybenzoic acid methyl ester) in 250 ml Ethanol]
5 ml Propionic acid

## Appendix II: Oligonucleotide sequences and their use

Primer name	Sequence (5' to 3')	Use
Hbs- <i>Taq</i> -1	CCTCTGAGCTCTCACTTTTGG	<i>Taq</i> -qPCR to test the expression of <i>hbs</i>
Hbs- <i>Taq</i> -2	GCACGACACTGATCTTAACCG	
CG30016- <i>Taq</i> -1	TTGGATACTTCGGTGGGAAAG	<i>Taq</i> -qPCR to test the expression of <i>CG30016</i>
CG30016- <i>Taq</i> -2	GGGATCTCCATTCTGAATCTC	
Uro- <i>Taq</i> -1	GACTTCAGCTCCATTGACAAC	<i>Taq</i> -qPCR to test the expression of <i>UO</i>
Uro- <i>Taq</i> -2	GAGACCCTTGATGCCCCG	
Uro-F-SgfI	TCAG GCGATCGCC ACC ATGTTTGCCACGCCCCCTC	Cloning for the expression in the cell-free expression system
Uro-R-PmeI-6His	ATGG GTTTAAAC CTAATGGTGATGGTGATGATG CAGGTGACTATTGATGTT	
CG30016-F-SgfI	GACC GCGATCGCC ACC ATGGATGCACGAAAGTTTTCT	
CG30016-R-PmeI-6His	GATG GTTTAAAC CTAATGGTGATGGTGATGATG TGTTCCACGATATGTGGA	
Uro-F-EcoRI	CGGGTACCACCATGTTTGCCACGCCCC TCAG	Cloning for the expression in <i>E. coli</i> cells
Uro-R-XbaI-6His	CGTCTAGACAGGTGACTATTGATGTTC C	
CG30016-F- EcoRI	CACCATGGATGCACGAAAGTTTTCTAC CCACATATTGGATACTTCGGTGGGAA AGG	
CG30016-R-XbaI-	CTAATGGTGATGGTGATGATGCTCGTT	

6His	CAGGTAGAACTGGCCTGTTCCACGAT ATGTGG	
Uro-F	CACCATGCATCATCACCATCACCATTTT GCCACGCCCCTCAGACAG	Cloning for the expression in S2 cells
Uro-R-6His	CTACAGGTGACTATTGATGTTCTTCCG GGCCAATTGGGCATAGATGGTGCCA	
CG30016-F	CACCATGGATGCACGAAAGTTTTCTAC CCACATATTGGATACTTCG	
CG30016-R-6His	CTAATGGTGATGGTGATGATGTGTCC ACGATATGTGGAATAC	

### Appendix III: Primary and secondary antibodies and their use

Antibody name	Type of antibody	Dilution	Use
Anti-6X His tag <sup>®</sup> antibody (abcam, UK)	Mouse monoclonal	1:1000	Western blotting – primary antibody
ECL <sup>™</sup> Anti-Mouse IgG, Horseradish Peroxidase-Linked (Amersham)	Sheep polyclonal	1:2500	Western blotting - secondary antibody
Anti-6X His tag <sup>®</sup> antibody (abcam, UK)	Mouse monoclonal	1:500	ICC using MTs – primary antibody
Alexa Fluor <sup>®</sup> 568-labelled anti-mouse IgG H & L (Molecular Probes)	Goat polyclonal	1:1000	ICC using MTs - secondary antibody
Anti-6X His tag <sup>®</sup> antibody (abcam, UK)	Mouse monoclonal	1:250	ICC using S2 cells – primary antibody
Alexa Fluor <sup>®</sup> 568-labelled anti-mouse IgG H & L (Molecular Probes)	Goat polyclonal	1:500	ICC using S2 cells - secondary antibody

## Appendix IV: *Drosophila* saline recipe

Compound	mM	g/l
Dissolve in around 800 ml of water while stirring:		
NaCl	117.5	6.86
KCl	20	1.49
CaCl <sub>2</sub> *2H <sub>2</sub> O	2	0.29
MgCl <sub>2</sub> *6H <sub>2</sub> O	2	0.41
Glucose	20	3.96
HEPES	8.6	2.05
Dissolve separately in 100 ml water each, then add to the rest of the mixture		
NaHCO <sub>3</sub>	10.24	0.86
NaH <sub>2</sub> PO <sub>4</sub>	4.5	0.70
Adjust pH to 6.8 and volume to 1 l, filter sterilise and store in a fridge		

## Appendix V: Standards used for metabolite identification at PF

Compound name	Retention time
Imidazole-4-acetate	17.77
N-Acetyl-D-glucosamine	11.42
1-Naphthylacetic acid	5.51
Melatonin	5.92
Phenylhydrazine	6.12

4-Aminobenzoate	6.24
Nicotinate	8.17
riboflavin	8.06
Glycerol	9.84
Inosine	10.28
L-Phenylalanine	13.12
L-Leucine	13.93
L-Tryptophan	14.05
2-Phenylglycine	13.93
Selenomethionine	14.22
L-Methionine	14.51
Guanine	15.09
Pyridoxine	16.66
L-Valine	15.51
Adenine	16.41
L-Proline	15.78
Serotonin	18.06
Taurine	15.45
trans-4-Hydroxy-L-proline	16.41
Creatinine	19
N2-Acetyl-L-lysine	18.89
L-Threonine	17.56

L-Aspartate	17.43
L-Glutamine	18.12
beta-Alanine	19.91
L-Asparagine	18.5
dGMP	14.82
L-Serine	18.62
L-Citrulline	19.12
Cytidine	20.53
Ethanolamine phosphate	18.44
Glycine	18.81
sn-glycero-3-Phosphocholine	18.89
L-Cystine	22.1
D-Glucosamine	24.04
meso-2,6-Diaminoheptanedioate	24.12
L-Histidine	26.72
L-2,4-Diaminobutanoate	27.53
L-Lysine	28.39
L-Arginine	28.2
Thiamin	33.58
4-Coumarate	5.91
(R)-3-Hydroxybutanoate	6.77
Phthalate	6.1

Fumarate	6.45
Methylmalonate	6.58
3,4-Dihydroxyphenylacetate	7.13
Thymidine	7.4
Phenolsulfonphthalein	4.8
Pyruvate	6.62
Malonate	7.34
(R)-2-Hydroxyglutarate	7.82
Deoxyuridine	8.09
pyrazinoate	8.28
Xanthine	9.02
Gallate	9.04
Uridine	9.29
cis-Aconitate	7.15
L-Rhamnose	10.21
Orotate	8.45
MOPS	13.3
D-Gluconic acid	13.44
Pyridoxal phosphate	13.4
D-Galacturonate	14.48
D-Galactarate	14.54
3',5'-Cyclic AMP	14



UMP	14.98
IMP	13.4
L-Cysteate	14.65
D-ribose 5-phosphate	15.65
CMP	18.51
D-glucose 6-phosphate	17.38
Phosphoenolpyruvate	18.59
6-Phospho-D-gluconate	22.28
L-Glutamate	17.28
D-glucose	13.94
Picolinic acid	9.25
Nicotinamide	10.72
5-Oxoproline	7.46
L-isoleucine	16.57
Pyridoxal	14.65
L-Tyrosine	14.92
4- Trimethylammoniobutanoate	17
Betaine	14.71
1-Aminocyclopropane-1- carboxylate	20.46
L-Alanine	19.34
L-homoserine	17.98

Choline	19.91
AMP	16.53
1-Aminopropan-2-ol	20.1
L-Ornithine	28.43
Phenylpyruvate	5.36
Mesaconate	6.47
succinate semialdehyde	6.28
Succinate	7.11
2-Oxoglutarate	7.3
N-Acetylglutamine	8.75
(S)-Malate	8.75
D-Ribose	12.89
N-Acetylneuraminate	12.76
L-Gulono-1,4-lactone	13.49
sucrose	17.11
D-Glucuronolactone	14.97
DL-Glyceraldehyde 3-phosphate	14.58
3-Phospho-D-glycerate	17.8
Isonicotinic acid	8.19
2-Aminobutan-4-olide	14.43
Benzoate	7.01
Acetoacetate	8.22

Itaconate	10.82
N-acetyl-L-glutamate	7.52
Maleic acid	8.2
D-Glucono-1,4-lactone	9.3
D-Fructose	15.89
Orthophosphate	14.86
lipoamide	5.9
Mercaptoethanol	8.56
Biopterin	14.22
L-Homocysteine	15.57
O-Acetyl-L-serine	14.43
L-cysteine	16.12
gamma-L-Glutamyl-L-cysteine	15.31
glutathione	15.92
acetylcysteine	6.5
ascorbate	8.2
Dihydrolipoamide	5.79
Homocystine	21.9
L-Cystine	22.07
L-Dehydroascorbate	8.2

## Appendix VI: All metabolites identified in *Drosophila* tissues and their relative abundance

Mass	RT	Formula	Putative metabolite	CNS	Gut	MTs	WF
195.09	5.79	C <sub>10</sub> H <sub>13</sub> NO <sub>3</sub>	Damascenine	148.92	106.28	70.73	0.00
244.10	5.99	C <sub>12</sub> H <sub>12</sub> N <sub>4</sub> O <sub>2</sub>	Flavin	0.00	0.00	63.84	0.00
193.06	8.08	C <sub>7</sub> H <sub>15</sub> NOS <sub>2</sub>	6-methylthiohexylhydroximate	161.15	0.00	55.44	0.00
118.03	6.84	C <sub>4</sub> H <sub>6</sub> O <sub>4</sub>	Succinate	4.79	63.15	34.16	1.00
129.04	7.64	C <sub>5</sub> H <sub>7</sub> NO <sub>3</sub>	L-1-Pyrroline-3-hydroxy-5-carboxylate	2.27	7.34	33.99	1.00
398.14	5.67	C <sub>22</sub> H <sub>22</sub> O <sub>7</sub>	Deoxypodophyllotoxin	0.00	0.00	23.30	0.00
114.03	6.35	C <sub>5</sub> H <sub>6</sub> O <sub>3</sub>	2-Hydroxy-2,4-pentadienoate	0.00	20.85	18.99	0.00
165.05	17.63	C <sub>5</sub> H <sub>11</sub> NO <sub>3</sub> S	L-Methionine S-oxide	0.00	0.27	16.81	1.00
175.06	6.02	C <sub>10</sub> H <sub>9</sub> NO <sub>2</sub>	Indole-3-acetate	35.87	7.25	15.23	0.00
241.11	7.24	C <sub>25</sub> H <sub>30</sub> N <sub>4</sub> O <sub>6</sub>	Phe-Gly-Pro-Tyr	3.12	15.15	14.45	1.00
148.05	5.79	C <sub>9</sub> H <sub>8</sub> O <sub>2</sub>	trans-Cinnamate	21.67	16.07	12.99	0.00
187.08	5.80	C <sub>8</sub> H <sub>13</sub> NO <sub>4</sub>	6-Acetamido-2-oxohexanoate	9.20	12.69	9.64	1.00
169.07	16.69	C <sub>8</sub> H <sub>11</sub> NO <sub>3</sub>	Pyridoxine	0.00	16.40	8.65	1.00
191.06	6.20	C <sub>7</sub> H <sub>13</sub> NO <sub>3</sub> S	N-Acetylmethionine	7.49	33.03	8.56	0.00
123.03	8.13	C <sub>6</sub> H <sub>5</sub> NO <sub>2</sub>	Nicotinate	10.49	30.04	7.65	1.00
758.51	4.78	C <sub>48</sub> H <sub>70</sub> O <sub>7</sub>	(3R,2'S)-Myxol 2'-(2,4-di-O-methyl- $\alpha$ -L-fucoside)	4.65	2.05	6.54	0.00

163.10	5.88	C10H13NO	N-Acetylphenylethylamine	20.05	13.74	6.53	1.00
175.03	5.86	C9H19O12P	nonulose 9-phosphate	10.67	8.88	6.06	1.00
236.08	13.17	C11H12N2O4	L-Formylkynurenine	0.00	0.30	5.48	1.00
190.08	5.81	C8H14O5	(R)-3-((R)-3-Hydroxybutanoyloxy)butanoate	8.71	5.32	5.45	1.00
109.05	15.31	C6H7NO	2-Aminophenol	0.62	0.18	5.28	1.00
208.08	13.46	C10H12N2O3	L-Kynurenine	0.00	0.03	5.27	1.00
173.07	6.06	C7H11NO4	N-Acetyl-L-glutamate 5-semialdehyde	4.64	9.41	4.93	1.00
191.04	8.87	C7H5N5O2	6-formyl-H2-pterin	0.04	0.13	4.59	1.00
145.07	6.89	C6H11NO3	4-Acetamidobutanoate	1.65	10.36	4.37	1.00
141.04	6.20	C6H7NO3	6-oxo-1,4,5,6-tetrahydronicotinate	7.10	9.04	4.17	1.00
138.04	12.81	C6H6N2O2	Urocanate	1.98	6.88	3.92	1.00
159.09	6.08	C7H13NO3	5-Acetamidopentanoate	2.22	4.06	3.91	1.00
173.14	14.98	C9H19NO2	Muscarine	0.00	1.41	3.76	1.00
113.05	6.08	C5H7NO2	1-Pyrroline-2-carboxylate	3.68	7.14	3.65	1.00
266.16	4.46	C12H26O4S	sodium dodecyl sulfate	3.86	3.18	3.00	1.00
376.14	8.00	C17H20N4O6	Riboflavin	0.00	0.13	2.91	1.00
115.03	6.23	C4H5NO3	Maleamate	1.39	2.76	2.48	1.00
297.09	11.04	C11H15N5O3S	5'-Methylthioadenosine	7.10	1.60	2.19	1.00
187.12	6.00	C9H17NO3	N-Heptanoylglycine	0.96	3.95	2.19	1.00

138.03	5.80	C7H6O3	4-Hydroxybenzoate	0.97	2.16	2.11	1.00
258.09	7.39	C10H14N2O6	Ribothymidine	0.00	0.59	2.10	1.00
242.08	5.98	C12H10N4O2	Lumichrome	0.63	0.27	2.00	1.00
122.05	9.63	C6H6N2O	Picolinamide	4.48	1.50	1.99	1.00
325.37	10.93	C22H47N	di-n-Undecylamine	1.73	1.86	1.94	1.00
146.06	6.30	C6H10O4	(S)-2-Aceto-2-hydroxybutanoate	2.32	8.71	1.88	1.00
154.03	6.06	C7H6O4	2,5-Dihydroxybenzoate	0.00	0.00	1.85	0.00
282.18	5.64	C16H26O4	12-trans-Hydroxy juvenile hormone III	2.70	1.86	1.83	1.00
179.04	10.64	C6H5N5O2	2-Amino-4,7-dihydroxypteridine	0.00	0.03	1.65	1.00
446.34	5.26	C28H46O4	3-Dehydroteasterone	3.95	3.07	1.65	1.00
216.10	7.23	C27H28O5	Aspulvinone H	1.39	2.18	1.62	1.00
168.07	13.58	C11H8N2	beta-Carboline	6.15	1.13	1.59	1.00
131.06	6.86	C5H9NO3	N-Acetyl-beta-alanine	1.23	3.54	1.58	1.00
270.18	5.64	C15H26O4	(10S)-Juvenile hormone III acid diol	2.61	1.73	1.54	1.00
124.06	7.13	C6H8N2O	Methylimidazole acetaldehyde	1.60	1.35	1.49	1.00
464.46	5.25	C31H60O2	Hentriacontane-14,16-dione	0.00	0.00	1.47	0.00
222.16	5.45	C14H22O2	Rishitin	2.43	1.53	1.41	1.00
194.06	5.78	C10H10O4	Ferulate	1.13	2.33	1.39	1.00
272.24	5.59	C16H32O3	16-hydroxypalmitate	4.05	2.07	1.33	1.00

112.03	7.99	C4H4N2O2	Uracil	2.03	2.09	1.23	1.00
166.06	5.91	C9H10O3	3-(3-Hydroxy-phenyl)-propanoic acid	1.87	3.31	1.23	1.00
99.03	6.26	C4H5NO2	Hymexazol	1.28	2.25	1.23	1.00
100.02	5.60	C4H4O3	2-oxobut-3-enanoate	1.39	1.03	1.21	1.00
201.14	5.77	C10H19NO3	Capryloylglycine	1.16	1.28	1.17	1.00
130.06	5.69	C6H10O3	(S)-3-Methyl-2-oxopentanoic acid	0.82	1.26	1.13	1.00
204.08	5.37	C12H12O3	3-Butylidene-7-hydroxyphthalide	1.74	1.56	1.12	1.00
422.23	4.34	C20H38O7S	1,4-Bis(2-ethylhexyl) sulfosuccinate	2.02	1.26	1.10	1.00
205.13	7.25	C9H19NO4	Pantothenol	0.49	1.74	1.05	1.00
178.06	5.77	C10H10O3	Coniferyl aldehyde	1.17	2.13	1.04	1.00
215.15	5.88	C11H21NO3	N-Nonanoylglycine	0.85	0.87	1.02	1.00
147.04	10.85	C5H9NO2S	Thiomorpholine 3-carboxylate	2.61	4.34	1.02	1.00
300.12	5.71	C14H20O7	Salidroside	0.69	0.31	0.99	1.00
139.03	5.72	C6H5NO3	6-Hydroxynicotinate	0.90	1.05	0.99	1.00
219.11	6.76	C9H17NO5	Pantothenate	2.17	17.72	0.98	1.00
328.12	6.04	C15H20O8	Paeonoside	0.87	1.12	0.97	1.00
158.06	5.71	C7H10O4	2-Isopropylmaleate	2.77	0.38	0.97	1.00
250.16	5.47	C15H22O3	Xanthoxin	0.93	1.40	0.95	1.00
408.38	5.31	C30H48	4,4'-Diapophytoene	2.48	2.96	0.94	1.00

229.89	5.42	C6H2OCl4	2,3,5,6-Tetrachlorophenol	4.77	1.18	0.94	1.00
131.09	5.78	C6H13NO2	1-nitrohexane	1.06	0.97	0.94	1.00
369.29	5.59	C21H39NO4	cis-5-Tetradecenoylcarnitine	1.06	1.05	0.94	1.00
164.05	5.82	C9H8O3	4-Coumarate	0.76	1.73	0.94	1.00
251.10	12.96	C10H13N5O3	Deoxyadenosine	5.21	1.18	0.94	1.00
70.08	5.60	C5H10	butylene	1.31	1.40	0.92	1.00
278.15	5.28	C16H22O4	2-Ethylhexyl phthalate	1.72	1.45	0.92	1.00
252.14	5.50	C14H20O4	ubiquinol-1	1.01	1.31	0.90	1.00
170.11	5.73	C13H14	Aethusin	1.41	0.91	0.88	1.00
184.07	5.83	C9H12O4	3-Methoxy-4-hydroxyphenylethyleneglycol	0.83	1.32	0.82	1.00
170.06	5.98	C8H10O4	3,4-Dihydroxyphenylethyleneglycol	1.21	1.89	0.81	1.00
220.11	5.42	C13H16O3	Precocene 2	1.00	0.82	0.80	1.00
110.04	6.22	C6H6O2	p-Benzenediol	0.77	0.82	0.80	1.00
99.07	7.31	C5H9NO	N-Methyl-2-pyrrolidinone	1.11	0.89	0.78	1.00
227.22	5.58	C14H29NO	myristic amide	1.24	0.76	0.75	1.00
243.18	5.75	C13H25NO3	N-Undecanoylglycine	0.91	0.96	0.75	1.00
161.05	8.92	C9H7NO2	4,6-Dihydroxyquinoline	0.07	0.02	0.75	1.00
316.17	5.39	C19H24O4	Gibberellin A9	1.00	1.24	0.74	1.00
264.17	5.53	C16H24O3	Dehydrojuvabione	1.28	0.70	0.73	1.00



120.06	5.65	C8H8O	Phenylacetaldehyde	0.81	0.83	0.71	1.00
326.28	5.43	C20H38O3	2-Oxophytanate	0.90	0.92	0.71	1.00
126.10	5.96	C8H14O	Sulcatone	0.95	0.80	0.71	1.00
229.17	5.82	C12H23NO3	N-Decanoylglycine	0.92	0.55	0.70	1.00
284.20	5.61	C16H28O4	(10S)-Juvenile hormone III diol	1.26	1.27	0.69	1.00
418.16	5.80	C22H26O8	Euparotin acetate	0.94	0.82	0.68	1.00
88.06	7.71	C3H8N2O	N,N'-Dimethylurea	1.31	0.84	0.68	1.00
268.08	10.27	C10H12N4O5	Inosine	0.00	1.05	0.67	1.00
386.17	5.65	C22H26O6	Burseran	0.89	0.95	0.67	1.00
129.12	5.90	C7H15NO	N-Methylhexanamide	0.98	0.96	0.67	1.00
240.07	5.26	C10H12N2O5	Dinoseb	0.70	0.71	0.66	1.00
204.12	5.62	C13H16O2	4'-Hydroxy-3'-prenylacetophenone	0.78	0.83	0.65	1.00
125.01	15.50	C2H7NO3S	Taurine	0.11	0.10	0.64	1.00
352.26	5.58	C21H36O4	Montanol	0.81	0.85	0.63	1.00
135.05	16.39	C5H5N5	Adenine	1.25	0.79	0.62	1.00
339.35	5.54	C22H45NO	Docosanamide	1.98	0.88	0.61	1.00
356.21	21.06	C16H28N4O5	Leu-Gln-Pro	0.91	0.73	0.58	1.00
268.17	5.56	C15H24O4	dihydroartemisinic acid hydroperoxide	0.63	0.55	0.57	1.00
136.05	5.69	C8H8O2	Phenylacetic acid	0.90	0.81	0.53	1.00

107.07	16.32	C7H9N	Benzylamine	0.82	0.71	0.53	1.00
236.18	5.56	C15H24O2	Capsidiol	0.80	0.70	0.52	1.00
152.03	9.06	C5H4N4O2	Xanthine	0.54	1.20	0.52	1.00
223.12	5.72	C12H17NO3	Cerulenin	1.16	0.22	0.51	1.00
194.13	5.63	C12H18O2	4-Hexyloxyphenol	0.72	0.71	0.51	1.00
244.07	9.25	C9H12N2O6	Uridine	0.11	1.86	0.51	1.00
255.26	5.59	C16H33NO	Palmiticamide	0.93	0.69	0.50	1.00
128.08	5.77	C7H12O2	3-Isopropylbut-3-enoic acid	0.86	0.88	0.50	1.00
271.21	5.66	C15H29NO3	Tridecanoylglycine	0.63	2.14	0.49	1.00
138.07	5.62	C8H10O2	4-Hydroxyphenylethanol	0.66	0.69	0.45	1.00
181.06	10.75	C6H7N5O2	8-Hydroxy-7-methylguanine	0.00	0.00	0.44	1.00
166.10	5.77	C10H14O2	Perillic acid	0.67	0.70	0.44	1.00
264.10	33.43	C12H16N4OS	Thiamin	0.00	0.39	0.43	1.00
165.07	15.25	C6H7N5O	3-Methylguanine	0.18	0.22	0.42	1.00
145.11	13.96	C7H15NO2	Acetylcholine	0.26	0.75	0.39	1.00
272.25	5.26	C20H32	Taxa-4(5),11(12)-diene	2.42	2.21	0.39	1.00
121.09	15.26	C8H11N	Phenethylamine	0.52	0.56	0.37	1.00
182.04	10.74	C6H6N4O3	1-Methyluric acid	0.00	0.02	0.37	1.00
252.09	9.29	C10H12N4O4	Deoxyinosine	0.00	0.07	0.33	1.00

180.11	5.69	C11H16O2	3-tert-Butyl-5-methylcatechol	0.46	0.42	0.33	1.00
297.11	9.73	C11H15N5O5	1-methylguanosine	0.00	0.00	0.30	1.00
105.08	20.59	C4H11NO2	Diethanolamine	0.48	0.35	0.29	1.00
122.07	5.68	C8H10O	Phenylethyl alcohol	0.37	0.36	0.28	1.00
192.03	5.65	C6H8O7	Carboxymethyloxysuccinate	1.00	0.59	0.28	1.00
151.05	9.73	C5H5N5O	2-Hydroxyadenine	0.00	0.00	0.28	1.00
322.21	5.54	C19H30O4	Decylubiquinone	0.22	0.24	0.27	1.00
296.20	5.62	C17H28O4	10-Deoxymethynolide	0.72	0.44	0.26	1.00
219.08	6.92	C21H26O10	Glaucolide B	0.02	0.00	0.25	1.00
136.04	9.65	C5H4N4O	Hypoxanthine	0.24	0.68	0.23	1.00
385.28	5.63	C21H39NO5	3-Hydroxy-cis-5-tetradecenoylcarnitine	0.35	20.58	0.21	1.00
78.01	8.55	C2H6OS	Mercaptoethanol	0.30	0.35	0.21	1.00
103.10	19.85	C5H13NO	Choline	0.01	0.41	0.19	1.00
85.05	7.50	C4H7NO	Acetone cyanohydrin	0.49	0.17	0.18	1.00
254.19	5.62	C15H26O3	3-hydroxy-15-dihydrolubimin	0.55	0.23	0.18	1.00
168.08	5.78	C9H12O3	1,3,5-trimethoxybenzene	0.28	0.32	0.18	1.00
150.05	10.32	C5H10O5	D-Xylulose	0.00	0.56	0.17	1.00
357.25	5.63	C19H35NO5	2,3-diocanoylglyceramide	0.12	23.41	0.15	1.00
257.20	5.63	C14H27NO3	N-Lauroylglycine	0.36	0.98	0.15	1.00

267.10	13.42	C10H13N5O4	Adenosine	0.20	0.51	0.14	1.00
270.05	5.68	C15H10O5	Apigenin	0.25	0.30	0.13	1.00
383.27	5.63	C21H37NO5	3-Hydroxy-5, 8-tetradecadiencarnitine	0.14	19.11	0.12	1.00
245.16	5.77	C12H23NO4	N-(octanoyl)-L-homoserine	0.00	0.23	0.10	1.00
240.15	16.76	C12H20N2O3	Slaframine	0.00	0.30	0.10	1.00
582.44	5.77	C41H58O2	Spheroidenone	0.00	2.58	0.06	1.00
329.22	5.67	C17H31NO5	6-Keto-decanoylcarnitine	0.05	28.88	0.05	1.00
152.05	5.73	C8H8O3	3,4-Dihydroxyphenylacetaldehyde	0.05	0.21	0.05	1.00
163.05	11.02	C6H5N5O	Pterin	0.04	0.01	0.03	1.00
168.04	5.93	C8H8O4	Homogentisate	0.04	0.28	0.02	1.00
169.09	15.84	C7H11N3O2	Nalpha-Methylhistidine	0.00	0.14	0.02	1.00
235.07	8.58	C9H9N5O3	6-Succinoaminopurine	0.00	0.00	0.02	1.00
427.37	11.33	C25H49NO4	Stearoylcarnitine	0.20	0.21	0.01	1.00
207.09	5.94	C11H13NO3	N-Acetyl-L-phenylalanine	0.00	27.14	0.00	0.00
244.02	7.33	C6H13O6PS	5-methylthiideoxyribose-1-phosphate	0.00	8.04	0.00	1.00
390.13	5.91	C20H22O8	Piceid	0.00	2.09	0.00	0.00
400.23	11.89	C18H32N4O6	Ala-Leu-Thr-Pro	0.00	1.98	0.00	1.00
473.21	11.89	C19H31N5O9	Glu-Thr-Gln-Pro	0.00	1.70	0.00	1.00
426.25	11.76	C20H34N4O6	Ile-Thr-Pro-Pro	0.00	1.62	0.00	1.00

393.23	5.80	C20H31N3O5	Leu-Val-Tyr	0.00	1.27	0.00	1.00
489.20	11.75	C21H27N7O7	Asn-Trp-Asn-Gly	0.00	1.23	0.00	1.00
537.22	11.76	C22H31N7O9	Asn-Asn-Gln-Tyr	0.00	1.21	0.00	1.00
360.11	6.52	C15H20O10	3-Methoxy-4-hydroxyphenylglycolglucuronide	0.00	1.00	0.00	1.00
372.16	5.60	C21H24O6	(-)-Arctigenin	0.65	0.83	0.00	1.00
394.42	5.27	C27H54O	2-heptacosanone	0.00	0.53	0.00	1.00
178.05	7.47	C6H10O6	2-Dehydro-3-deoxy-D-gluconate	0.00	0.46	0.00	1.00
126.04	7.18	C5H6N2O2	Thymine	0.00	0.45	0.00	1.00
794.62	5.29	C54H82O4	Ubiquinone-9	0.10	0.40	0.00	1.00
414.35	5.26	C28H46O2	22alpha-Hydroxy-campest-4-en-3-one	0.26	0.34	0.00	1.00
423.33	11.39	C25H45NO4	Linoelaidylcarnitine	0.03	0.33	0.00	1.00
425.35	11.36	C25H47NO4	Elaidiccarnitine	0.08	0.25	0.00	1.00
455.40	11.25	C27H53NO4	Arachidylcarnitine	0.07	0.19	0.00	1.00
726.56	5.22	C49H74O4	[PR] Coenzyme Q8	0.00	0.18	0.00	1.00
776.69	5.34	C15H11I4NO4	Thyroxine	0.08	0.14	0.00	1.00
143.04	8.33	C6H9NOS	5-(2-Hydroxyethyl)-4-methylthiazole	93.05	0.00	0.00	0.00
210.08	5.82	C13H10N2O	2-Aminoacridone	61.63	0.00	0.00	0.00
167.06	5.72	C8H9NO3	3-Methoxyanthranilate	36.98	0.00	0.00	0.00
163.09	14.07	C7H9N5	1-ethyladenine	28.24	0.00	0.00	0.00

164.03	7.71	C6H4N4O2	2,4-Dihydroxypteridine	0.30	0.00	0.00	1.00
142.06	5.90	C7H10O3	4-Oxocyclohexanecarboxylate	0.00	0.00	0.00	1.00
109.03	5.76	C6H5O2	Benzosemiquinone	0.00	0.00	0.00	1.00
95.04	16.08	C5H5NO	3-hydroxypyridine	0.00	0.00	0.00	1.00
284.18	17.00	C13H24N4O3	Melanostatin	0.00	0.00	0.00	1.00
209.11	6.13	C11H15NO3	p-Lactophenetide	0.00	0.00	0.00	1.00
282.10	8.30	C11H14N4O5	1-Methylinosine	0.00	0.00	0.00	1.00
198.05	5.85	C9H10O5	3-(3,4-Dihydroxyphenyl)lactate	0.00	0.00	0.00	1.00
139.07	18.81	C6H9N3O	L-Histidinal	0.00	0.00	0.00	1.00
191.06	6.03	C10H9NO3	5-Hydroxyindoleacetate	0.00	0.00	0.00	1.00
448.36	5.25	C28H48O4	Teasterone	0.00	0.00	0.00	1.00
210.09	5.69	C11H14O4	Sinapyl alcohol	0.00	0.00	0.00	1.00
224.07	5.59	C11H12O5	Sinapate	0.00	0.00	0.00	1.00
195.08	10.72	C7H9N5O2	2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine	0.00	0.00	0.00	1.00
243.12	18.29	C10H17N3O4	Ala-Gly-Pro	0.00	0.00	0.00	1.00

## Appendix VII: All metabolites identified in *Drosophila* whole fly *CG30016* mutant and the control

Mass	RT	Formula	Putative metabolite	WFLY17767	WFLY18554
205.04	7.79	C10H7NO4	Xanthurenic acid	1.00	28.31
383.14	15.89	C14H25NO11	N-Acetyllactosamine	1.00	21.07
182.08	13.85	C6H14O6	D-Sorbitol	1.00	6.38
218.05	13.88	C13H21N6O9P	(L-Seryl)adenylate	1.00	5.13
126.04	15.31	C5H6N2O2	Imidazole-4-acetate	1.00	4.08
182.04	10.84	C6H6N4O3	1-Methyluric acid	1.00	3.52
284.08	11.78	C12H16N2O4S	Cys-Tyr	1.00	3.44
212.01	18.95	C5H9O7P	P-DPD	1.00	3.42
283.09	12.39	C10H13N5O5	Guanosine	1.00	3.31
163.05	10.68	C6H5N5O	Pterin	1.00	3.10
233.13	15.00	C10H19NO5	Hydroxypropionylcarnitine	1.00	3.10
111.99	16.08	CH5O4P	Hydroxymethylphosphonate	1.00	3.07
208.05	22.30	C6H12N2O4S	Cys-Ser	1.00	2.99
175.10	18.80	C6H13N3O3	L-Citrulline	1.00	2.96
141.02	19.70	C2H8NO4P	Ethanolamine phosphate	1.00	2.95
268.08	10.29	C10H12N4O5	Inosine	1.00	2.58

110.05	14.95	C5H6N2O	Imidazole-4-acetaldehyde	1.00	2.52
297.05	24.33	C8H15N3O5S2	L-Cysteinylglycinedisulfide	1.00	2.30
145.11	14.49	C7H15NO2	4-Trimethylammoniobutanoate	1.00	2.28
136.04	9.26	C5H4N4O	allopurinol	1.00	2.19
152.03	8.66	C5H4N4O2	Xanthine	1.00	2.01
148.04	7.54	C5H8O5	(R)-2-Hydroxyglutarate	1.00	1.97
278.09	13.14	C10H18N2O5S	Glu-Met	1.00	1.96
132.09	26.33	C5H12N2O2	L-Ornithine	1.00	1.93
134.06	9.41	C5H10O4	Deoxyribose	1.00	1.92
344.33	5.05	C21H44O3	1-O-Octadecyl-sn-glycerol	1.00	1.84
342.12	15.81	C12H22O11	Sucrose	1.00	1.82
198.07	12.66	C17H24N4O5S	Phe-Cys-Gln	1.00	1.81
123.03	7.27	C6H5NO2	Nicotinate	1.00	1.78
188.15	26.88	C9H20N2O2	N6,N6,N6-Trimethyl-L-lysine	1.00	1.77
244.07	8.98	C9H12N2O6	Uridine	1.00	1.77
161.07	13.83	C6H11NO4	L-2-Aminoadipate	1.00	1.70
240.02	22.27	C6H12N2O4S2	L-Cystine	1.00	1.68
243.09	18.84	C9H13N3O5	Cytidine	1.00	1.64
477.31	9.37	C21H43N5O7	Gentamicin	1.00	1.61



255.10	11.53	C9H13N5O4	2-Amino-4-hydroxy-6-(D-erythro-1,2,3-trihydroxypropyl)-7,8-dihydropteridine	1.00	1.61
151.58	23.91	C11H21N5O5	Glu-Arg	1.00	1.56
129.09	15.14	C5H11N3O	4-Guanidinobutanal	1.00	1.53
120.09	4.97	C9H12	1,2,4-Trimethylbenzene	1.00	1.47
212.12	14.89	C10H16N2O3	Pro-Pro	1.00	1.47
446.06	21.40	C11H20N4O11P2	CDP-ethanolamine	1.00	1.46
264.10	28.46	C12H16N4OS	Thiamin	1.00	1.45
165.05	17.81	C5H11NO3S	L-Methionine S-oxide	1.00	1.42
206.09	19.65	C7H14N2O5	Thr-Ser	1.00	1.39
183.07	23.33	C5H14NO4P	Choline phosphate	1.00	1.37
168.03	11.77	C5H4N4O3	Urate	1.00	1.35
108.57	28.99	C9H19N3O3	gamma-L-Glutamylputrescine	1.00	1.32
188.03	6.41	C7H8O6	(Z)-But-2-ene-1,2,3-tricarboxylate	1.00	1.32
232.11	15.94	C9H16N2O5	N2-Succinyl-L-ornithine	1.00	1.32
224.08	5.02	C15H12O2	Flavanone	1.00	1.30
357.14	8.68	C16H23NO8	Bakankoside	1.00	1.29
186.10	15.05	C8H14N2O3	Ala-Pro	1.00	1.28
217.11	19.40	C8H15N3O4	Ala-Ala-Gly	1.00	1.27
190.06	19.46	C6H10N2O5	Asp-Gly	1.00	1.26

259.05	17.90	C6H14NO8P	alpha-D-Glucosamine 1-phosphate	1.00	1.26
222.07	22.44	C7H14N2O4S	L-Cystathionine	1.00	1.24
173.07	14.67	C7H11NO4	N-Acetyl-L-glutamate 5-semialdehyde	1.00	1.24
126.00	11.41	C2H6O4S	2-Hydroxyethanesulfonate	1.00	1.21
204.07	18.78	C7H12N2O5	Glu-Gly	1.00	1.20
142.07	14.93	C6H10N2O2	Ectoine	1.00	1.18
130.03	9.66	C5H6O4	2,5-Dioxopentanoate	1.00	1.18
203.09	18.65	C7H13N3O4	Ala-Asn	1.00	1.17
123.58	29.23	C10H21N3O4	Lys-Thr	1.00	1.16
234.08	19.51	C8H14N2O6	Glu-Ser	1.00	1.16
509.17	24.74	C17H31N7O7S2	Glu-Cys-Cys-Arg	1.00	1.16
116.57	29.65	C9H19N3O4	Lys-Ser	1.00	1.15
175.55	21.73	C12H21N3O7S	Met-Asp-Ser	1.00	1.15
101.12	14.20	C6H15N	Hexylamine	1.00	1.13
101.56	29.08	C8H17N3O3	Lys-Gly	1.00	1.12
195.09	8.71	C10H13NO3	L-Tyrosine methyl ester	1.00	1.10
89.05	17.74	C3H7NO2	L-Alanine	1.00	1.10
145.16	43.25	C7H19N3	Spermidine	1.00	1.09
178.05	9.66	C6H10O6	D-Glucono-1,5-lactone	1.00	1.08

75.03	18.26	C2H5NO2	Glycine	1.00	1.08
230.16	12.23	C11H22N2O3	Leu-Val	1.00	1.08
593.22	17.56	C26H35N5O9S	Glu-Glu-Met-Trp	1.00	1.07
88.02	6.93	C3H4O3	Pyruvate	1.00	1.07
155.13	13.06	C9H17NO	N-Methylpelletierine	1.00	1.06
129.59	25.92	C12H25N3O3	Leu-Lys	1.00	1.06
386.17	5.18	C22H26O6	Burseran	1.00	1.05
136.59	27.36	C11H23N5O3	Val-Arg	1.00	1.05
135.05	14.60	C5H5N5	Adenine	1.00	1.05
187.05	6.71	C7H9NO5	2-(Acetamidomethylene)succinate	1.00	1.02
260.03	19.45	C6H13O9P	D-Glucose 6-phosphate	1.00	1.02
275.11	17.75	C10H17N3O6	Gamma-Glutamylglutamine	1.00	1.02
132.05	18.35	C4H8N2O3	L-Asparagine	1.00	1.01
175.05	7.24	C6H9NO5	N-Acetyl-L-aspartate	1.00	1.01
155.07	24.75	C6H9N3O2	L-Histidine	1.00	0.99
363.06	18.59	C10H14N5O8P	GMP	1.00	0.99
192.10	15.03	C17H28N4O6	Ala-Thr-Pro-Pro	1.00	0.99
122.05	8.37	C6H6N2O	Nicotinamide	1.00	0.98
115.06	15.38	C5H9NO2	L-Proline	1.00	0.98

202.13	13.40	C9H18N2O3	Leu-Ala	1.00	0.97
181.18	10.50	C12H23N	Dicyclohexylamine	1.00	0.97
195.08	10.48	C7H9N5O2	2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine	1.00	0.97
138.07	21.46	C11H20N2O6	N6-(L-1,3-Dicarboxypropyl)-L-lysine	1.00	0.96
103.10	17.24	C5H13NO	Choline	1.00	0.96
105.04	18.38	C3H7NO3	L-Serine	1.00	0.95
129.04	6.93	C5H7NO3	L-1-Pyrroline-3-hydroxy-5-carboxylate	1.00	0.95
218.13	14.69	C9H18N2O4	Leu-Ser	1.00	0.95
154.00	14.92	C3H7O5P	Propanoyl phosphate	1.00	0.95
179.09	11.81	C10H13NO2	(-)-Salsolinol	1.00	0.94
81.98	15.10	H3O3P	Phosphonate	1.00	0.94
702.54	5.81	C50H70O2	demethylmenaquinone-8	1.00	0.94
146.11	26.04	C6H14N2O2	L-Lysine	1.00	0.94
289.13	15.16	C11H19N3O6	Val-Asp-Gly	1.00	0.94
252.11	14.24	C12H16N2O4	Ala-Tyr	1.00	0.94
165.08	12.19	C9H11NO2	L-Phenylalanine	1.00	0.94
131.09	12.87	C6H13NO2	L-Leucine	1.00	0.93
138.03	5.70	C7H6O3	4-Hydroxybenzoate	1.00	0.92
446.19	25.72	C20H26N6O6	Gln-Tyr-His	1.00	0.92

103.06	16.66	C4H9NO2	N,N-Dimethylglycine	1.00	0.92
111.08	31.22	C5H9N3	1H-Imidazole-4-ethanamine	1.00	0.92
229.11	14.49	C9H15N3O4	Asn-Pro	1.00	0.91
423.33	9.78	C25H45NO4	Linoelaidylcarnitine	1.00	0.91
220.11	18.83	C8H16N2O5	Thr-Thr	1.00	0.91
118.03	6.55	C4H6O4	Succinate	1.00	0.91
146.06	6.17	C6H10O4	(S)-2-Aceto-2-hydroxybutanoate	1.00	0.91
345.19	13.33	C15H27N3O6	Leu-Val-Asp	1.00	0.90
345.05	15.05	C10H12N5O7P	3',5'-Cyclic GMP	1.00	0.90
219.11	6.85	C9H17NO5	Pantothenate	1.00	0.90
279.12	14.20	C13H17N3O4	Phe-Asn	1.00	0.90
305.04	18.78	C9H12N3O7P	2',3'-Cyclic CMP	1.00	0.90
147.05	16.65	C5H9NO4	L-Glutamate	1.00	0.90
204.09	12.96	C11H12N2O2	L-Tryptophan	1.00	0.88
201.17	14.92	C11H23NO2	Caproylcholine	1.00	0.88
133.04	17.38	C4H7NO4	L-Aspartate	1.00	0.88
294.06	25.75	C9H15N2O7P	Pyrimidine 5'-nucleotide	1.00	0.88
225.11	14.62	C20H30N6O6	His-Thr-Pro-Pro	1.00	0.88
369.04	12.21	C9H15N5O7S2	Amidosulfuron	1.00	0.87

216.12	16.19	C8H16N4O3	N-acetyl-(L)-arginine	1.00	0.87
327.22	12.26	C16H29N3O4	Leu-Val-Pro	1.00	0.87
149.05	13.73	C5H11NO2S	L-Methionine	1.00	0.87
133.07	15.74	C5H11NO3	DL-&beta;-hydroxynorvaline	1.00	0.87
244.18	11.61	C12H24N2O3	Leucyl-leucine	1.00	0.87
239.10	12.84	C9H13N5O3	Dihydrobiopterin	1.00	0.87
246.12	14.73	C10H18N2O5	Glu-Val	1.00	0.86
197.07	17.02	C9H11NO4	3,4-Dihydroxy-L-phenylalanine	1.00	0.86
235.07	8.25	C9H9N5O3	6-Succinoaminopurine	1.00	0.86
329.05	14.91	C10H12N5O6P	3',5'-Cyclic AMP	1.00	0.85
467.13	16.34	C19H25N5O5S2	Cys-Trp-Cys-Gly	1.00	0.85
171.03	19.55	C3H10NO5P	serinol phosphate	1.00	0.85
146.07	17.89	C5H10N2O3	L-Glutamine	1.00	0.85
187.17	30.27	C9H21N3O	N1-Acetylspermidine	1.00	0.85
245.15	29.91	C9H19N5O3	beta-Alanyl-L-arginine	1.00	0.84
264.15	11.97	C14H20N2O3	Phe-Val	1.00	0.84
285.17	13.80	C13H23N3O4	Leu-Gly-Pro	1.00	0.83
262.13	12.28	C14H18N2O3	Phe-Pro	1.00	0.83
347.06	17.88	C10H14N5O7P	dGMP	1.00	0.83

341.23	11.72	C17H31N3O4	Leu-Leu-Pro	1.00	0.83
228.04	18.32	C14H24N4O7S3	Glu-Cys-Cys-Cys	1.00	0.83
100.06	14.95	C4H8N2O	Gyromitrin	1.00	0.82
224.08	14.11	C10H12N2O4	3-Hydroxy-L-kynurenine	1.00	0.82
208.08	12.44	C10H12N2O3	L-Kynurenine	1.00	0.81
267.10	12.27	C10H13N5O4	Adenosine	1.00	0.81
169.07	14.86	C8H11NO3	Pyridoxine	1.00	0.80
158.04	12.90	C4H6N4O3	Allantoin	1.00	0.80
169.08	24.82	C7H11N3O2	N(pi)-Methyl-L-histidine	1.00	0.79
181.07	14.56	C9H11NO3	3-Amino-3-(4-hydroxyphenyl)propanoate	1.00	0.78
189.04	6.54	C10H7NO3	Kynurenate	1.00	0.78
216.15	13.06	C10H20N2O3	Val-Val	1.00	0.78
172.08	14.78	C7H12N2O3	Glycylproline	1.00	0.77
151.03	17.07	C4H9NO3S	methiin	1.00	0.75
278.16	11.38	C15H22N2O3	Leu-Phe	1.00	0.75
189.06	6.75	C7H11NO5	Glutaryl glycine	1.00	0.75
290.12	23.57	C10H18N4O6	N-(L-Arginino)succinate	1.00	0.75
237.09	11.74	C9H19NO2S2	8-[(aminomethyl)sulfanyl]-6-sulfanyloctanoic acid	1.00	0.74
281.11	19.05	C25H34N6O7S	Glu-Met-Phe-His	1.00	0.74

240.15	14.79	C12H20N2O3	Slaframine	1.00	0.73
311.18	14.22	C15H25N3O4	Val-Pro-Pro	1.00	0.72
116.07	27.73	C10H20N2O4	Spermic acid 2	1.00	0.71
297.09	9.60	C11H15N5O3S	5'-Methylthioadenosine	1.00	0.71
116.01	6.11	C4H4O4	Fumarate	1.00	0.70
247.14	14.35	C11H21NO5	Hydroxybutyrylcarnitine	1.00	0.69
152.06	14.82	C7H8N2O2	N1-Methyl-2-pyridone-5-carboxamide	1.00	0.69
190.11	26.55	C6H14N4O3	N-(omega)-Hydroxyarginine	1.00	0.69
181.06	10.77	C6H7N5O2	8-Hydroxy-7-methylguanine	1.00	0.66
145.09	14.98	C5H11N3O2	4-Guanidinobutanoate	1.00	0.66
431.16	16.59	C16H25N5O9	Asn-Asp-Pro-Ser	1.00	0.66
137.08	14.26	C8H11NO	Tyramine	1.00	0.66
175.08	14.96	C7H13NO4	alpha-aminopimelate	1.00	0.65
214.13	14.41	C10H18N2O3	Val-Pro	1.00	0.65
220.08	12.29	C11H12N2O3	5-Hydroxy-L-tryptophan	1.00	0.64
244.15	14.62	C20H40N8O6	Arg-Lys-Val-Ser	1.00	0.64
489.17	16.58	C14H30N6O11P	O-1,4-alpha-L-Dihydrostreptosyl-streptidine 6-phosphate	1.00	0.63
137.05	15.30	C7H7NO2	N-Methylnicotinate	1.00	0.63
157.11	11.07	C8H15NO2	Homostachydrine	1.00	0.62



310.12	14.22	C14H18N2O6	Glu-Tyr	1.00	0.62
238.00	9.21	C12H8OCl2	2,6-Dichloro-4'-biphenylol	1.00	0.61
306.08	21.15	C20H32N6O12S2	Glutathione disulfide	1.00	0.59
250.06	14.87	C8H14N2O5S	gamma-L-Glutamyl-L-cysteine	1.00	0.57
152.05	5.32	C8H8O3	3,4-Dihydroxyphenylacetaldehyde	1.00	0.56
176.04	10.44	C5H8N2O5	N-Carbamoyl-L-aspartate	1.00	0.54
167.11	14.80	C8H13N3O	2-Dimethylamino-5,6-dimethylpyrimidin-4-ol	1.00	0.53
213.04	21.61	C13H22N4O8S2	Asp-Cys-Cys-Ser	1.00	0.52
461.17	17.09	C17H27N5O10	Asn-Val-Asp-Asp	1.00	0.52
265.12	19.28	C25H34N6O5S	His-Met-Phe-Pro	1.00	0.51
461.17	17.08	C21H27N5O5S	Cys-Trp-Gly-Pro	1.00	0.50
270.13	14.71	C22H36N8O8	Glu-Lys-Gln-His	1.00	0.48
308.16	26.76	C12H24N2O7	Fructoselysine	1.00	0.40
504.17	16.45	C18H32O16	Maltotriose	1.00	0.39
227.09	14.30	C9H13N3O4	Deoxycytidine	1.00	0.38
324.04	15.70	C9H13N2O9P	Pseudouridine 5'-phosphate	1.00	0.38
189.08	7.93	C8H15NO2S	Prenyl-L-cysteine	1.00	0.31
156.02	9.95	C5H4N2O4	Orotate	1.00	0.29
160.12	25.55	C7H16N2O2	N6-Methyl-L-lysine	1.00	0.27

161.05	10.38	C6H11NO2S	allylcysteine	1.00	0.25
121.02	15.72	C3H7NO2S	D-Cysteine	1.00	0.10
407.27	4.98	C23H37NO5	Norerythrostachaldine	1.00	0.09
307.08	15.50	C10H17N3O6S	Ala-Asp-Cys	1.00	0.09

# Appendix VIII: All metabolites identified in *Drosophila* adult MTs CG30016 mutant and the control

Mass	RT	Formula	Putative metabolite	MT17767	MT18554
161.11	16.64	C7H15NO3	L-Carnitine	1.00	2.98
125.01	16.28	C2H7NO3S	Taurine	1.00	1.62
268.08	10.27	C10H12N4O5	Inosine	1.00	1.32
141.02	19.98	C2H8NO4P	Ethanolamine phosphate	1.00	1.31
267.10	12.30	C10H13N5O4	Adenosine	1.00	1.29
244.07	8.96	C9H12N2O6	Uridine	1.00	1.26
103.10	17.61	C5H13NO	Choline	1.00	1.19
149.05	13.68	C5H11NO2S	L-Methionine	1.00	1.12
257.10	20.08	C8H20NO6P	sn-glycero-3-Phosphocholine	1.00	1.08
132.05	18.40	C4H8N2O3	L-Asparagine	1.00	1.06
155.07	25.03	C6H9N3O2	L-Histidine	1.00	1.04
165.08	12.19	C9H11NO2	L-Phenylalanine	1.00	0.99
146.07	17.94	C5H10N2O3	L-Glutamine	1.00	0.98
204.09	12.92	C11H12N2O2	L-Tryptophan	1.00	0.98
119.06	17.26	C4H9NO3	L-Threonine	1.00	0.97
130.03	6.22	C5H6O4	Itaconate	1.00	0.97

131.09	12.87	C6H13NO2	L-Leucine	1.00	0.97
118.03	6.77	C4H6O4	Succinate	1.00	0.97
133.04	17.40	C4H7NO4	L-Aspartate	1.00	0.95
147.05	16.82	C5H9NO4	L-Glutamate	1.00	0.94
135.05	14.68	C5H5N5	Adenine	1.00	0.88
75.03	18.30	C2H5NO2	Glycine	1.00	0.87
89.05	16.95	C3H7NO2	L-Alanine	1.00	0.86
105.04	18.44	C3H7NO3	L-Serine	1.00	0.81
190.10	23.49	C7H14N2O4	meso-2,6-Diaminoheptanedioate	1.00	0.76
88.10	31.94	C4H12N2	Putrescine	1.00	0.66
152.03	8.65	C5H4N4O2	Xanthine	1.00	0.62
202.03	10.49	C5H6N4O5	5-Hydroxy-2-oxo-4-ureido-2,5-dihydro-1H-imidazole-5-carboxylate	1.00	5.75
217.13	12.83	C10H19NO4	O-Propanoylcarnitine	1.00	4.51
260.03	19.44	C6H13O9P	D-Glucose-6-phosphate	1.00	3.38
425.35	9.78	C25H47NO4	Elaidiccarnitine	1.00	3.26
427.37	9.73	C25H49NO4	Stearoylcarnitine	1.00	3.10
187.12	5.73	C9H17NO3	N-Heptanoylglycine	1.00	2.98
242.08	5.58	C12H10N4O2	Lumichrome	1.00	2.82
306.08	21.20	C20H32N6O12S2	Glutathione disulfide	1.00	2.31

168.03	11.78	C5H4N4O3	Urate	1.00	2.25
731.55	7.64	C40H78NO8P	[PC (14:0/18:1)] 1-tetradecanoyl-2-(11Z-octadecenoyl)-sn-glycero-3-phosphocholine	1.00	2.01
729.53	7.63	C40H76NO8P	[PC (14:0/18:2)] 1-tetradecanoyl-2-(9Z,12Z-octadecadienoyl)-sn-glycero-3-phosphocholine	1.00	1.97
689.50	5.67	C37H72NO8P	PC(14:1(9Z)/15:0)	1.00	1.94
687.48	5.67	C37H70NO8P	[PE (16:1/16:1)] 1,2-di-(9Z-hexadecenoyl)-sn-glycero-3-phosphoethanolamine	1.00	1.91
158.04	12.90	C4H6N4O3	Allantoin	1.00	1.83
649.47	7.83	C34H68NO8P	[PE (14:0/15:0)] 1-tetradecanoyl-2-pentadecanoyl-sn-glycero-3-phosphoethanolamine	1.00	1.74
685.47	5.67	C37H68NO8P	PE(14:0/18:3(6Z,9Z,12Z))	1.00	1.73
455.40	9.64	C27H53NO4	Arachidylcarnitine	1.00	1.59
228.15	14.84	C11H20N2O3	Leu-Pro	1.00	1.55
184.02	11.88	C5H4N4O4	5-Hydroxyisourate	1.00	1.52
250.06	15.01	C8H14N2O5S	gamma-L-Glutamyl-L-cysteine	1.00	1.50
283.09	12.43	C10H13N5O5	Guanosine	1.00	1.48
537.51	5.03	C34H67NO3	[SP (16:0)] N-(hexadecanoyl)-sphing-4-enine	1.00	1.45
208.08	12.43	C10H12N2O3	L-Kynurenine	1.00	1.44
565.54	5.02	C36H71NO3	[SP (18:0)] N-(octadecanoyl)-sphing-4-enine	1.00	1.41
467.13	16.39	C19H25N5O5S2	Cys-Trp-Cys-Gly	1.00	1.38
771.58	7.51	C43H82NO8P	[PE (18:0/20:2)] 1-octadecanoyl-2-(11Z,14Z-eicosadienoyl)-sn-	1.00	1.37

			glycero-3-phosphoethanolamine		
213.04	21.65	C13H22N4O8S2	Asp-Cys-Cys-Ser	1.00	1.37
307.08	15.69	C10H17N3O6S	Ala-Asp-Cys	1.00	1.37
137.08	14.27	C8H11NO	Tyramine	1.00	1.36
111.08	31.29	C5H9N3	1H-Imidazole-4-ethanamine	1.00	1.36
165.05	17.87	C5H11NO3S	L-Methionine S-oxide	1.00	1.35
394.32	5.08	C28H42O	Ergosta-5,7,22,24(28)-tetraen-3beta-ol	1.00	1.35
769.56	7.54	C43H80NO8P	PC(15:0/20:3(5Z,8Z,11Z))	1.00	1.35
259.05	17.86	C6H14NO8P	alpha-D-Glucosamine 1-phosphate	1.00	1.34
876.55	4.60	C44H80N2O15	Megalomicin A	1.00	1.31
331.19	5.06	C20H27O4	7'-carboxy-gama-tocotrienol	1.00	1.31
351.31	5.07	C22H41NO2	[FA (20:2)] N-(11Z,14Z-eicosadienoyl)-ethanolamine	1.00	1.30
528.45	4.95	C35H60O3	[PR] 32,35-anhydrobacteriohopaneterol	1.00	1.30
224.08	5.03	C15H12O2	Flavanone	1.00	1.30
757.56	7.54	C42H80NO8P	[PC (16:0/18:2)] 1-hexadecanoyl-2-(9Z,12Z-octadecadienoyl)-sn-glycero-3-phosphocholine	1.00	1.29
105.08	16.04	C4H11NO2	Diethanolamine	1.00	1.27
744.49	4.41	C40H73O10P	1-18:2-2-trans-16:1-phosphatidylglycerol	1.00	1.27
163.08	15.82	C6H13NO4	D-perosamine	1.00	1.22
215.06	17.98	C5H14NO6P	sn-glycero-3-Phosphoethanolamine	1.00	1.22

713.50	5.64	C39H72NO8P	[PE (16:0/18:3)] 1-hexadecanoyl-2-(9Z,12Z,15Z-octadecatrienoyl)-sn-glycero-3-phosphoethanolamine	1.00	1.20
216.12	16.75	C8H16N4O3	N-acetyl-(L)-arginine	1.00	1.20
148.04	7.76	C5H8O5	(R)-2-Hydroxylglutarate	1.00	1.19
464.17	5.41	C23H28O10	Enhydrin	1.00	1.16
755.55	7.58	C42H78NO8P	[PC (16:0/18:3)] 1-hexadecanoyl-2-(9Z,12Z,15Z-octadecatrienoyl)-sn-glycero-3-phosphocholine	1.00	1.15
426.31	5.10	C28H42O3	[ST (5:0/3:0)] (5Z,7E,22E,24E)-(1S,3R)-24a-homo-9,10-seco-5,7,10(19),22,24-cholestapentaene-1,3,25-triol	1.00	1.15
208.09	5.05	C15H12O	Chalcone	1.00	1.14
244.15	14.63	C20H40N8O6	Arg-Lys-Val-Ser	1.00	1.14
155.13	13.02	C9H17NO	N-Methylpelletierine	1.00	1.12
223.01	17.37	C6H10NO4PS	4-Methyl-5-(2-phosphoethyl)-thiazole	1.00	1.11
216.04	14.68	C5H13O7P	2-C-Methyl-D-erythritol 4-phosphate	1.00	1.10
186.14	11.71	C9H18N2O2	N-(3-acetamidopropyl)-4-aminobutanal	1.00	1.08
353.33	5.02	C22H43NO2	[FA (20:0)] N-(11Z-eicosaenoyl)-ethanolamine	1.00	1.07
345.05	14.78	C10H12N5O7P	3',5'-Cyclic GMP	1.00	1.05
104.01	7.98	C3H4O4	Hydroxypyruvate	1.00	1.04
181.04	15.82	C5H11NO4S	DL-Methionine sulfone	1.00	1.04
324.11	16.47	C12H20O10	Bis-D-fructose 2',1:2,1'-dianhydride	1.00	1.04
248.05	12.91	C9H12O8	pentane-1,3,4,5-tetracarboxylate	1.00	1.04

136.02	7.48	C4H8O3S	S-Methyl-1-thio-D-glycerate	1.00	1.03
146.02	9.35	C5H6O5	Methyloxaloacetate	1.00	1.03
201.17	15.19	C11H23NO2	Caproylcholine	1.00	1.03
308.16	26.79	C12H24N2O7	Fructoselysine	1.00	1.03
181.07	14.62	C9H11NO3	3-Amino-3-(4-hydroxyphenyl)propanoate	1.00	1.02
121.09	13.09	C8H11N	Phenethylamine	1.00	1.02
428.33	5.04	C28H44O3	Ercalcitriol	1.00	1.02
139.14	11.96	C9H17N	Pinidine	1.00	0.99
422.06	16.26	C15H20NO9S2	Gluconasturtiin	1.00	0.98
737.50	5.65	C41H72NO8P	PE(14:0/22:5(4Z,7Z,10Z,13Z,16Z))	1.00	0.98
139.98	13.69	C2H4O5S	Sulfoacetate	1.00	0.98
81.98	14.55	H3O3P	Phosphonate	1.00	0.98
107.07	14.05	C7H9N	N-Methylaniline	1.00	0.97
243.22	10.83	C14H29NO2	[SP (14:0)] tetradecasphing-4E-ene	1.00	0.97
216.10	7.23	C27H28O5	Aspulvinone H	1.00	0.97
181.18	10.52	C12H23N	Dicyclohexylamine	1.00	0.96
442.34	5.05	C29H46O3	4alpha-Methylzymosterol-4-carboxylate	1.00	0.96
231.06	16.46	C22H22O11	Isoscoparine	1.00	0.95
739.51	5.59	C41H74NO8P	[PE (16:0/20:4)] 1-hexadecanoyl-2-(5Z,8Z,11Z,14Z-eicosatetraenoyl)-sn-glycero-3-phosphoethanolamine	1.00	0.94



101.12	14.19	C6H15N	Hexylamine	1.00	0.93
446.06	26.01	C11H20N4O11P2	CDP-ethanolamine	1.00	0.92
444.32	5.05	C28H44O4	[ST (4:0/4:0)] (5Z,7E,22E)-(1S,3R,24R)-9,10-seco-5,7,10(19),22-ergostatetraene-1,3,24,25-tetrol	1.00	0.90
238.10	5.05	C16H14O2	[Fv Methox] 4'-Methoxychalcone	1.00	0.90
183.07	24.23	C5H14NO4P	Choline phosphate	1.00	0.89
129.12	14.40	C7H15NO	4-Trimethylammoniobutanal	1.00	0.88
243.18	5.35	C13H25NO3	N-Undecanoylglycine	1.00	0.88
145.07	26.81	C6H11NO3	5-hydroxy-pipecolate	1.00	0.87
71.07	14.61	C4H9N	3-Buten-1-amine	1.00	0.86
688.55	7.41	C38H77N2O6P	[SP (18:0/14:0)] N-(octadecanoyl)-tetradecasphing-4-enine-1-phosphoethanolamine	1.00	0.86
386.30	5.05	C22H42O5	1,2-dioctanoyl-1,2,6-hexanetriol	1.00	0.85
160.11	5.40	C8H16O3	Ethyl (R)-3-hydroxyhexanoate	1.00	0.84
253.01	11.70	C10H17N6O12P3	AMPPNP	1.00	0.83
329.05	14.44	C10H12N5O6P	3',5'-Cyclic dGMP	1.00	0.83
117.12	13.16	C6H15NO	2-Methylcholine	1.00	0.83
781.56	7.54	C44H80NO8P	[PC (16:0/20:4)] 1-hexadecanoyl-2-(5Z,8Z,11Z,14Z-eicosatetraenoyl)-sn-glycero-3-phosphocholine	1.00	0.82
285.27	10.03	C17H35NO2	[SP methyl(16:0)] 15-methyl-hexadecasphing-4E-enine	1.00	0.82
795.58	7.52	C45H82NO8P	PE(18:0/22:4(7Z,10Z,13Z,16Z))	1.00	0.81

194.06	5.58	C10H10O4	Ferulate	1.00	0.79
87.10	14.99	C5H13N	3-Methylbutanamine	1.00	0.77
172.07	5.57	C8H12O4	[FA dioxo(8:0)] 4,7-dioxo-octanoic acid	1.00	0.76
138.07	21.49	C11H20N2O6	N6-(L-1,3-Dicarboxypropyl)-L-lysine	1.00	0.76
284.23	5.03	C17H32O3	[FA methoxy(16:1)] 2-methoxy-5Z-hexadecenoic acid	1.00	0.75
175.08	15.26	C7H13NO4	alpha-aminopimelate	1.00	0.74
254.08	23.77	C6H15N4O5P	L-Arginine phosphate	1.00	0.73
208.15	5.17	C13H20O2	4-Heptyloxyphenol	1.00	0.73
369.32	5.07	C22H43NO3	[SP amino,tetramethyl(4:0/18:0/3:0)] 2S-amino-5,9,13,17-tetramethyl-8E,16-octadecadiene-1,3R,14-triol	1.00	0.71
132.04	5.95	C5H8O4	2-Acetolactate	1.00	0.70
453.29	7.46	C21H44NO7P	[PE (16:0)] 1-hexadecanoyl-sn-glycero-3-phosphoethanolamine	1.00	0.67
240.15	14.92	C12H20N2O3	Slaframine	1.00	0.64
145.08	15.28	C5H11N3O2	4-Guanidinobutanoate	1.00	0.63
179.06	5.72	C9H9NO3	Hippurate	1.00	0.54
160.07	5.64	C7H12O4	[FA (7:0/2:0)] Heptanedioic acid	1.00	0.48
146.06	5.83	C6H10O4	(S)-2-Aceto-2-hydroxybutanoate	1.00	0.45
174.09	5.49	C8H14O4	Suberic acid	1.00	0.41
161.07	15.97	C6H11NO4	L-2-Aminoadipate	1.00	0.41
175.10	18.88	C6H13N3O3	L-Citruline	1.00	0.37

129.15	10.91	C <sub>8</sub> H <sub>19</sub> N	Octylamine	1.00	0.32
--------	-------	----------------------------------	------------	------	------

## Appendix IX: All metabolites identified in *Drosophila melanogaster* secreted fluid

Mass	RT	Formula	Putative metabolite	Control	Sample
208.08	11.74	C <sub>10</sub> H <sub>12</sub> N <sub>2</sub> O <sub>3</sub>	L-Kynurenine	0.00	697.73
224.08	14.28	C <sub>10</sub> H <sub>12</sub> N <sub>2</sub> O <sub>4</sub>	3-Hydroxy-L-kynurenine	0.00	196.37
168.03	14.29	C <sub>5</sub> H <sub>4</sub> N <sub>4</sub> O <sub>3</sub>	Urate	1.00	156.89
205.04	12.73	C <sub>10</sub> H <sub>7</sub> NO <sub>4</sub>	Xanthurenic acid	0.00	125.81
216.12	16.47	C <sub>8</sub> H <sub>16</sub> N <sub>4</sub> O <sub>3</sub>	N-acetyl-(L)-arginine	1.00	97.35
136.02	10.95	C <sub>4</sub> H <sub>8</sub> O <sub>3</sub> S	S-Methyl-1-thio-D-glycerate	0.00	83.87
283.09	13.95	C <sub>10</sub> H <sub>13</sub> N <sub>5</sub> O <sub>5</sub>	Guanosine	0.00	74.66
260.03	18.02	C <sub>6</sub> H <sub>13</sub> O <sub>9</sub> P	D-Glucose-6-Phosphate	0.00	71.29
191.06	11.74	C <sub>10</sub> H <sub>9</sub> NO <sub>3</sub>	5-Hydroxyindoleacetate	0.00	63.71
129.06	7.35	C <sub>9</sub> H <sub>7</sub> N	Isoquinoline	0.00	61.82
398.14	17.70	C <sub>15</sub> H <sub>22</sub> N <sub>6</sub> O <sub>5</sub> S	S-Adenosyl-L-methionine	0.00	56.43
268.08	12.01	C <sub>10</sub> H <sub>12</sub> N <sub>4</sub> O <sub>5</sub>	Inosine	1.00	49.25
416.10	12.71	C <sub>15</sub> H <sub>20</sub> N <sub>4</sub> O <sub>8</sub> S	O-Carbamoyl-deacetylcephalosporin C	0.00	41.72
163.10	4.52	C <sub>10</sub> H <sub>13</sub> NO	(R)-2-Methylimino-1-phenylpropan-1-ol	1.00	36.87
347.06	14.52	C <sub>10</sub> H <sub>14</sub> N <sub>5</sub> O <sub>7</sub> P	3'-AMP	1.00	31.57
145.07	7.34	C <sub>6</sub> H <sub>11</sub> NO <sub>3</sub>	4-Acetamidobutanoate	1.00	26.63
198.07	9.57	C <sub>13</sub> H <sub>24</sub> N <sub>4</sub> O <sub>8</sub> S	Cys-Thr-Ser-Ser	0.00	23.01
248.05	16.75	C <sub>16</sub> H <sub>26</sub> N <sub>4</sub> O <sub>8</sub> P <sub>2</sub> S	2-Methyl-1-hydroxypropyl-ThPP	0.00	21.72

109.05	14.38	C6H7NO	2-Aminophenol	0.00	20.77
186.06	11.84	C7H10N2O4	(S)-AMPA	0.00	19.58
334.07	18.17	C9H19O11P	sn-glycero-3-Phospho-1-inositol	0.00	19.36
175.03	8.68	C9H19O12P	nonulose 9-phosphate	0.00	18.73
278.07	12.83	C9H15N2O6P	Pyrimidine 5'-deoxynucleotide	0.00	18.39
293.14	15.55	C14H19N3O4	Phe-Ala-Gly	0.00	18.20
158.04	15.35	C4H6N4O3	(S)(+)-Allantoin	0.00	17.45
236.08	10.71	C11H12N2O4	L-Formylkynurenine	1.00	14.79
79.97	13.13	[PO3]3-	Phosphite	0.00	14.54
143.06	10.17	C6H9NO3	Vinylacetyl glycine	1.00	14.02
183.07	16.85	C5H14NO4P	Choline phosphate	1.00	13.97
221.09	9.16	C9H19NOS2	8-methylthiooctylhydroximate	0.00	13.39
235.12	4.27	C13H17NO3	Lophophorine	1.00	12.13
113.05	10.82	C5H7NO2	(S)-1-Pyrroline-5-carboxylate	1.00	11.96
377.06	11.92	C13H18N2O7PS	external aldimine	0.00	11.49
161.07	16.81	C6H11NO4	L-2-Aminoai pate	0.00	10.85
223.03	9.49	C10H9NO3S	2-naphthylamine-1-sulfonate	0.00	10.35
72.02	10.86	C3H4O2	Methylglyoxal	0.00	10.24
692.11	17.89	C20H26N10O14P2	Guanosinediphosphateadenosine	0.00	9.68
278.09	13.78	C10H18N2O5S	Glu-Met	0.00	9.46
244.05	12.19	C9H12N2O4S	penem CGP31608	0.00	9.46
383.11	16.47	C14H17N5O8	Succinyladenosine	1.00	8.61

249.09	7.69	C10H19NO2S2	S-Acetyldihydrolipoamide-E	0.00	8.58
212.08	8.73	C18H24N4O8	Asp-Gln-Tyr	1.00	8.21
162.09	6.04	C7H14O4	beta-Cymaropyranose	1.00	7.81
311.13	8.64	C17H17N3O3	Imazaquin	1.00	7.16
146.02	14.96	C4H6N2O2S	ZAPA	1.00	6.42
179.09	13.51	C10H13NO2	(-)-Salsolinol	0.00	6.08
117.06	7.96	C8H7N	Indole	0.00	6.06
205.07	4.71	C11H11NO3	Gentianamine	0.00	6.06
139.03	14.59	C6H5NO3	6-Hydroxynicotinate	0.00	6.05
276.08	10.82	C10H16N2O5S	Biotinsulfone	0.00	5.92
427.03	17.25	C10H15N5O10P2	ADP	0.00	5.33
120.06	4.77	C8H8O	Phenylacetaldehyde	1.00	5.15
111.98	8.36	CH4O4S	Monomethyl sulfate	1.00	4.89
179.08	8.25	C7H9N5O	7-Aminomethyl-7-carbaguanine	0.00	4.73
243.09	13.10	C9H13N3O5	Cytidine	1.00	4.65
176.06	14.26	C9H8N2O2	4-Hydroxyaminoquinoline N-oxide	0.00	4.57
168.04	12.01	C8H8O4	Homogentisate	0.00	4.15
147.07	4.75	C9H9NO	3-Methyloxindole	0.00	3.98
97.98	17.96	H3O4P	Orthophosphate	1.00	3.92
207.09	4.81	C11H13NO3	Cantleyine	1.00	3.76
185.11	4.88	C9H15NO3	8-keto-7-aminoperlagonate	1.00	3.71
202.07	11.09	C11H10N2O2	alpha,beta-Didehydrotryptophan	1.00	3.71

214.10	7.79	C9H14N2O4	Pyrimidine nucleoside	1.00	3.52
206.04	19.76	C7H10O7	2-Hydroxybutane-1,2,4-tricarboxylate	1.00	3.18
155.09	4.84	C8H13NO2	Arecoline	1.00	3.18
144.04	11.11	C6H8O4	2,3-Dimethylmaleate	0.00	3.17
363.06	17.77	C10H14N5O8P	GMP	1.00	3.13
165.05	14.83	C5H11NO3S	L-Methionine S-oxide	1.00	3.04
138.04	7.32	C6H6N2O2	Urocanate	1.00	3.02
161.11	14.79	C7H15NO3	L-Carnitine	1.00	2.96
192.03	20.34	C6H8O7	Citrate	1.00	2.89
374.12	7.37	C16H22O10	Secologanate	1.00	2.86
173.11	5.34	C8H15NO3	Hexanoylglycine	1.00	2.84
125.06	11.60	C5H7N3O	5-Methylcytosine	0.00	2.55
151.05	13.64	C5H5N5O	Guanine	1.00	2.52
231.11	12.21	C10H17NO5	Suberylglycine	1.00	2.39
184.02	14.12	C5H4N4O4	5-Hydroxyisourate	0.00	2.38
223.08	8.31	C11H13NO4	Bendiocarb	1.00	2.35
284.08	13.75	C10H12N4O6	Xanthosine	1.00	2.20
156.02	11.21	C5H4N2O4	Orotate	0.00	2.19
138.03	4.76	C7H6O3	Gentisate aldehyde	1.00	2.18
170.09	3.95	C9H14O3	Furfural diethyl acetal	1.00	2.07
149.11	9.23	C6H15NO3	Triethanolamine	1.00	2.01
215.15	4.24	C11H21NO3	N-Nonanoylglycine	1.00	2.01

257.10	12.31	C10H15N3O5	5-Methylcytidine	1.00	1.82
169.07	8.06	C8H11NO3	Pyridoxine	1.00	1.81
209.07	7.25	C7H15NO4S	MOPS	1.00	1.74
134.06	9.32	C5H10O4	Deoxyribose	1.00	1.74
160.11	5.00	C8H16O3	Ethyl (R)-3-hydroxyhexanoate	1.00	1.65
243.18	4.38	C13H25NO3	N-Undecanoylglycine	1.00	1.65
104.01	17.91	C3H4O4	Malonate	1.00	1.61
200.08	10.31	C8H12N2O4	Dihydroclavaminic acid	1.00	1.61
149.08	4.65	C9H11NO	D-Cathinone	1.00	1.60
200.18	3.78	C12H24O2	Dodecanoic acid	1.00	1.60
105.08	11.04	C4H11NO2	Diethanolamine	1.00	1.59
123.03	7.47	C6H5NO2	Nicotinate	1.00	1.52
157.04	12.91	C6H7NO4	2-Aminomuconate	1.00	1.44
127.03	11.86	C5H5NO3	2,3,6-Trihydroxypyridine	1.00	1.43
166.10	4.97	C10H14O2	Perillic acid	1.00	1.38
168.08	4.84	C9H12O3	1,3,5-trimethoxybenzene	1.00	1.37
145.04	11.83	C5H7NO4	2-Oxoglutaramate	1.00	1.31
229.17	4.58	C12H23NO3	N-Decanoylglycine	1.00	1.29
174.10	7.87	C7H14N2O3	Val-Gly	1.00	1.27
238.13	6.85	C24H36N4O6	Ile-Phe-Thr-Pro	1.00	1.24
228.21	3.69	C14H28O2	Tetradecanoic acid	1.00	1.24
172.15	4.04	C10H20O2	Decanoic acid	1.00	1.23



160.04	16.71	C6H8O5	2-Oxoadipate	1.00	1.23
184.11	4.05	C10H16O3	4,5-dihydro-5,5-dimethyl-4-(3-oxobutyl)furan-2(3H)-one	1.00	1.23
376.14	8.85	C17H20N4O6	Riboflavin	1.00	1.21
182.08	15.51	C6H14O6	D-Sorbitol	1.00	1.20
236.14	3.93	C14H20O3	4-Heptyloxybenzoic acid	1.00	1.20
264.17	3.99	C16H24O3	Dehydrojuvabione	1.00	1.17
110.05	13.06	C5H6N2O	Imidazole-4-acetaldehyde	1.00	1.16
215.06	17.34	C5H14NO6P	sn-glycero-3-Phosphoethanolamine	1.00	1.15
284.27	3.59	C18H36O2	Octadecanoic acid	1.00	1.14
473.28	7.30	C21H39N5O7	Gln-Leu-Leu-Thr	1.00	1.14
174.06	12.89	C6H10N2O4	N-Formimino-L-glutamate	1.00	1.13
430.16	4.02	C23H26O8	Sesartemin	1.00	1.12
260.05	12.82	C9H12N2O5S	2-thiouridine	1.00	1.09
252.17	3.90	C15H24O3	Juvenile hormone III acid	1.00	1.09
266.06	12.80	C10H10N4O5	5'-Oxoinosine	1.00	1.09
128.12	3.97	C8H16O	1-Octanal	1.00	1.08
100.05	4.01	C5H8O2	Methyl methacrylate	1.00	1.08
197.07	11.64	C9H11NO4	3,4-Dihydroxy-L-phenylalanine	1.00	1.08
250.19	3.98	C16H26O2	Methyl farnesoate	1.00	1.06
130.10	4.38	C7H14O2	Ethyl isovalerate	1.00	1.04
162.07	4.13	C10H10O2	Methyl cinnamate	1.00	1.01
208.15	3.89	C13H20O2	4-Heptyloxyphenol	1.00	1.00

308.11	4.15	C20H12N4	porphyrin-ring	1.00	1.00
255.26	4.02	C16H33NO	Palmiticamide	1.00	0.99
254.22	3.83	C16H30O2	(9Z)-Hexadecenoic acid	1.00	0.99
339.35	2.62	C22H45NO	Docosanamide	1.00	0.97
220.08	11.73	C11H12N2O3	5-Hydroxy-L-tryptophan	1.00	0.96
94.04	9.73	C6H6O	Arene oxide	1.00	0.95
72.06	4.53	C4H8O	Butanal	1.00	0.95
196.06	15.95	C6H12O7	D-Gluconic acid	1.00	0.94
238.10	11.71	C8H18N2O4S	HEPES	1.00	0.93
116.08	4.71	C6H12O2	4-Hydroxyhexan-3-one	1.00	0.92
201.14	4.40	C10H19NO3	Capryloylglycine	1.00	0.86
150.07	4.75	C9H10O2	4-Coumaryl alcohol	1.00	0.86
179.08	12.35	C6H13NO5	D-Galactosamine	1.00	0.85
329.05	9.38	C10H12N5O6P	2',3'-Cyclic AMP	1.00	0.84
236.13	7.68	C11H16N4O2	CPX	1.00	0.82
324.04	16.09	C9H13N2O9P	UMP	1.00	0.81
118.08	4.00	C9H10	alpha-Methylstyrene	1.00	0.81
178.11	9.75	C10H14N2O	(S)-6-Hydroxynicotine	1.00	0.77
150.05	13.55	C5H10O5	D-Xylulose	1.00	0.72
166.03	16.02	C8H6O4	Phthlate	1.00	0.72
194.06	4.73	C10H10O4	5-Hydroxyconiferaldehyde	1.00	0.72
175.08	13.05	C7H13NO4	Calystegin B2	1.00	0.68

206.11	3.91	C16H14	trans,trans-1,4-Diphenyl-1,3-butadiene	1.00	0.68
230.15	3.89	C12H22O4	Diisopropyl adipate	1.00	0.67
194.08	12.86	C7H14O6	1-O-Methyl-myo-inositol	1.00	0.63
175.10	17.65	C6H13N3O3	L-Citruline	1.00	0.61
114.08	10.71	C5H10N2O	L-proline amide	1.00	0.61
180.05	15.27	C8H8N2O3	Nicotinurate	1.00	0.60
97.97	20.16	H2O4S	Sulfate	1.00	0.59
397.11	12.66	C15H19N5O6S	S-Adenosyl-4-methylthio-2-oxobutanoate	1.00	0.56
244.11	17.18	C10H16N2O5	Glu-Pro	1.00	0.55
208.10	10.13	C17H28N4O8	Asp-Val-Pro-Ser	1.00	0.55
409.15	10.80	C18H23N3O8	Glu-Asp-Phe	1.00	0.54
459.26	3.97	C26H37NO6	Militarinone A	1.00	0.53
384.12	14.97	C14H20N6O5S	S-Adenosyl-L-homocysteine	1.00	0.53
200.98	19.20	C3H7NO5S2	S-Sulfo-L-cysteine	1.00	0.52
187.06	12.72	C11H9NO2	Indoleacrylicacid	1.00	0.52
80.96	20.15	HO3S	HSO3-	1.00	0.48
474.13	12.71	C16H23N6O9P	glutamyl-beta-ketophosphonate-adenosine	1.00	0.47
308.16	22.03	C12H24N2O7	Fructoselysine	1.00	0.44
173.08	15.89	C6H11N3O3	5-Guanidino-2-oxopentanoate	1.00	0.42
204.09	12.72	C11H12N2O2	L-Tryptophan	1.00	0.41
128.06	16.47	C5H8N2O2	5,6-Dihydrothymine	1.00	0.41
381.24	4.74	C18H31N5O4	Leu-Leu-His	1.00	0.40

418.34	3.61	C27H46O3	7alpha,12alpha-Dihydroxy-5beta-cholestan-3-one	1.00	0.40
305.04	13.98	C9H12N3O7P	2',3'-Cyclic CMP	1.00	0.36
352.06	14.13	C12H17O10P	Arbutin 6-phosphate	1.00	0.35
264.10	21.51	C12H16N4OS	Thiamin	1.00	0.34
85.05	11.37	C4H7NO	Acetone cyanohydrin	1.00	0.34
115.06	14.22	C5H9NO2	L-Proline	1.00	0.33
163.08	12.59	C6H13NO4	1-deoxynojirimycin	1.00	0.33
360.14	12.54	C12H20N6O7	Asn-Asn-Asn	1.00	0.31
122.05	7.34	C6H6N2O	Picolinamide	1.00	0.31
306.03	11.98	C9H11N2O8P	2',3'-Cyclic UMP	1.00	0.31
178.05	16.34	C6H10O6	D-Glucono-1,5-lactone	1.00	0.31
256.11	9.57	C11H16N2O5	1-(beta-D-Ribofuranosyl)-1,4-dihydronicotinamide	1.00	0.30
159.09	18.33	C7H13NO3	3-Dehydrocarnitine	1.00	0.30
281.11	7.44	C11H15N5O4	1-Methyladenosine	1.00	0.30
202.14	22.37	C8H18N4O2	NG,NG-Dimethyl-L-arginine	1.00	0.29
133.02	8.67	C4H7NO2S	L-thiazolidine-4-carboxylate	1.00	0.28
191.03	17.26	C6H9NO4S	a Cysteine adduct	1.00	0.26
236.12	4.75	C12H16N2O3	Carbetamide	1.00	0.26
224.09	12.17	C9H12N4O3	Temurin	1.00	0.25
103.10	20.99	C5H13NO	Choline	1.00	0.23
210.07	15.66	C7H14O7	Sedoheptulose	1.00	0.23
309.11	16.93	C11H19NO9	O-Acetylneuraminic acid	1.00	0.23

132.05	16.92	C4H8N2O3	L-Asparagine	1.00	0.23
258.09	11.38	C10H14N2O6	(1-Ribosylimidazole)-4-acetate	1.00	0.22
149.05	12.62	C5H11NO2S	L-Methionine	1.00	0.22
136.04	14.09	C4H8O5	[FA trihydroxy(4:0)] 2,3,4-trihydroxy-butanoic acid	1.00	0.21
186.10	15.69	C8H14N2O3	Ala-Pro	1.00	0.20
231.13	18.56	C8H17N5O3	Gly-Arg	1.00	0.20
134.02	18.19	C4H6O5	(S)-Malate	1.00	0.20
133.04	17.01	C4H7NO4	L-Aspartate	1.00	0.20
165.08	10.88	C9H11NO2	L-Phenylalanine	1.00	0.19
135.05	9.56	C5H5N5	Adenine	1.00	0.19
247.07	18.30	C9H13NO7	N-Succinyl-L-glutamate	1.00	0.18
484.06	17.94	C14H22N4O9P2S	alpha,beta-Dihydroxyethyl-TPP	1.00	0.18
370.07	9.45	C16H20O6P2	BPH-674	1.00	0.18
267.10	9.43	C10H13N5O4	Adenosine	1.00	0.17
205.09	15.40	C8H15NO5	N-Acetyl-D-fucosamine	1.00	0.17
219.11	9.20	C9H17NO5	Pantothenate	1.00	0.17
155.07	16.25	C6H9N3O2	L-Histidine	1.00	0.16
174.11	26.87	C6H14N4O2	L-Arginine	1.00	0.16
131.09	11.78	C6H13NO2	L-Leucine	1.00	0.15
478.18	9.42	C17H30N6O8S	Asn-Met-Gln-Ser	1.00	0.15
158.07	17.67	C6H10N2O3	5-Hydroxyectoine	1.00	0.15
251.10	8.27	C10H13N5O3	Deoxyadenosine	1.00	0.14

280.11	12.77	C13H16N2O5	Phe-Asp	1.00	0.14
144.09	25.50	C6H12N2O2	L-isoglutamine	1.00	0.14
87.03	17.77	C3H5NO2	2-Aminoacrylate	1.00	0.14
146.07	16.72	C5H10N2O3	L-Glutamine	1.00	0.13
147.09	11.50	C6H13NO3	Fagomine	1.00	0.13
148.04	17.29	C5H8O5	D-Xylonolactone	1.00	0.13
323.07	16.87	C20H34N6O12S3	bisorganyltrisulfane	1.00	0.12
229.06	14.59	C9H11NO6	4-(L-Alanin-3-yl)-2-hydroxy-cis,cis-muconate 6-semialdehyde	1.00	0.11
196.08	15.52	C18H24N4O6	Asn-Pro-Tyr	1.00	0.11
174.14	22.17	C8H18N2O2	Ne,Ne dimethyllysine	1.00	0.11
112.03	8.64	C4H4N2O2	Uracil	1.00	0.10
105.04	17.46	C3H7NO3	L-Serine	1.00	0.10
188.15	23.12	C9H20N2O2	N6,N6,N6-Trimethyl-L-lysine	1.00	0.10
251.13	24.41	C24H34N6O6	Ala-Leu-Tyr-His	1.00	0.10
223.11	15.70	C8H17NO6	N-acetyl -D- glucosaminitol	1.00	0.10
219.07	17.27	C8H13NO6	O-Succinyl-L-homoserine	1.00	0.10
171.05	21.22	C7H9NO4	2,3,4,5-Tetrahydrodipicolinate	1.00	0.09
172.01	16.17	C3H9O6P	sn-Glycerol 3-phosphate	1.00	0.09
102.03	17.52	C4H6O3	2-Oxobutanoate	1.00	0.09
249.03	21.21	C8H11NO6S	Norepinephrinesulfate	1.00	0.09
529.09	16.63	C16H25N3O13P2	dTDP-3-amino-2,3,6-trideoxy-D-threo-hexopyranos-4-ulose	1.00	0.09
183.02	17.44	C4H9NO5S	L-Homocysteic acid	1.00	0.09

354.06	12.56	C15H14O10	2-Caffeoylisocitrate	1.00	0.09
75.03	17.38	C2H5NO2	Glycine	1.00	0.09
189.06	16.12	C7H11NO5	N-Acetyl-L-glutamate	1.00	0.09
89.05	17.07	C3H7NO2	L-Alanine	1.00	0.09
203.13	18.37	C8H17N3O3	Lys-Gly	1.00	0.08
181.07	14.39	C9H11NO3	L-Tyrosine	1.00	0.08
294.14	19.57	C27H40N8O5S	Arg-Met-Trp-Pro	1.00	0.08
286.08	14.14	C16H14O5	(S)-DNPA	1.00	0.07
147.05	16.70	C5H9NO4	L-Glutamate	1.00	0.07
256.14	12.43	C23H40N6O5S	His-Leu-Leu-Met	1.00	0.07
261.12	12.14	C11H19NO6	Lotaustralin	1.00	0.07
275.15	18.69	C11H21N3O5	L-a-glutamyl-L-Lysine	1.00	0.07
146.02	17.67	C5H6O5	2-Oxoglutarate	1.00	0.07
257.14	16.93	C11H19N3O4	Ala-Ala-Pro	1.00	0.07
504.17	19.07	C18H32O16	Maltotriose	1.00	0.06
119.06	16.06	C4H9NO3	L-Threonine	1.00	0.06
251.10	15.82	C9H17NO7	Muramic acid	1.00	0.06
345.05	14.30	C10H12N5O7P	3',5'-Cyclic GMP	1.00	0.06
303.15	18.58	C11H21N5O5	Glu-Arg	1.00	0.06
240.02	18.00	C6H12N2O4S2	L-Cystine	1.00	0.06
332.05	16.66	C16H12O8	3,3',4',5,7-Pentahydroxy-8-methoxyflavone	1.00	0.06
114.03	16.50	C5H6O3	2-Hydroxy-2,4-pentadienoate	1.00	0.05

121.02	18.01	C3H7NO2S	D-Cysteine	1.00	0.05
324.10	16.55	C12H20O10	Bis-D-fructose 2',1:2,1'-dianhydride	1.00	0.05
118.03	17.19	C4H6O4	Succinate	1.00	0.05
257.10	16.17	C8H20NO6P	sn-glycero-3-Phosphocholine	1.00	0.05
100.02	17.32	C4H4O3	2-oxobut-3-enanoate	1.00	0.05
171.06	18.16	C6H9N3O3	6-diazo-5-oxonorleucine	1.00	0.04
128.09	17.84	C6H12N2O	L-Lysine 1,6-lactam	1.00	0.04
260.15	18.42	C10H20N4O4	Lys-Asn	1.00	0.04
426.09	18.99	C13H22N4O8S2	S-glutathionyl-L-cysteine	1.00	0.04
96.02	17.86	C5H4O2	Furfural	1.00	0.03
289.14	18.86	C10H19N5O5	Asp-Arg	1.00	0.03
353.14	12.66	C16H23N3O4S	Met-Phe-Gly	1.00	0.03
152.03	12.74	C5H4N4O2	Xanthine	1.00	0.02
291.10	14.24	C11H17NO8	2,7-Anhydro-alpha-N-acetylneuraminic acid	1.00	0.02
146.11	25.50	C6H14N2O2	L-Lysine	1.00	0.02
335.13	16.24	C12H21N3O8	N4-(Acetyl-beta-D-glucosaminyl)asparagine	1.00	0.02
275.03	18.14	C6H15NO7P2	Phosphophosphinate	1.00	0.01
180.06	17.87	C6H12O6	myo-Inositol	1.00	0.01
84.02	17.92	C4H4O2	3-Butynoate	1.00	0.01
358.08	16.69	C14H18N2O7S	Miraxanthin-I	1.00	0.01
342.12	17.85	C12H22O11	Maltose	1.00	0.01
447.16	17.54	C16H25N5O10	Ala-Asp-Asp-Gln	1.00	0.01



246.12	14.21	C10H18N2O5	Glu-Val	1.00	0.00
245.14	9.30	C10H19N3O4	Leu-Asn	1.00	0.00
498.20	14.12	C26H30N2O8	strictosamide	1.00	0.00
380.04	25.52	C10H13N4O10P	Urate D-ribonucleotide	1.00	0.00
492.13	26.81	C17H24N4O13	Glu-Asp-Asp-Asp	1.00	0.00
415.28	12.66	C19H37N5O5	Ala-Lys-Val-Val	1.00	0.00
188.12	7.31	C8H16N2O3	Glycyl-leucine	1.00	0.00
217.14	17.89	C9H19N3O3	gamma-L-Glutamylputrescine	1.00	0.00
360.24	14.24	C16H32N4O5	Leu-Lys-Thr	1.00	0.00
366.06	16.26	C10H15N4O9P	1-(5'-Phosphoribosyl)-5-formamido-4-imidazolecarboxamide	1.00	0.00
402.14	17.85	C15H22N4O9	Asp-Asp-Gly-Pro	1.00	0.00
303.16	4.88	C16H21N3O3	Trp-Val	1.00	0.00
238.10	13.47	C11H14N2O4	Gly-Tyr	1.00	0.00
273.18	15.85	C11H23N5O3	Val-Arg	1.00	0.00
218.09	17.04	C8H14N2O5	L-Ala-L-Glu	1.00	0.00
520.23	25.50	C23H32N6O8	His-Thr-Thr-Tyr	1.00	0.00
353.07	17.44	C9H16N5O8P	2,5-Diamino-6-(5'-phosphoribosylamino)-4-pyrimidineone	1.00	0.00
344.22	14.10	C14H28N6O4	Leu-Gly-Arg	1.00	0.00
217.11	13.88	C8H15N3O4	N-Acetyl-L-citrulline	1.00	0.00
317.17	4.74	C17H23N3O3	Leu-Trp	1.00	0.00
245.15	17.80	C9H19N5O3	beta-Alanyl-L-arginine	1.00	0.00
289.13	15.28	C11H19N3O6	Pro-Ser-Ser	1.00	0.00

243.16	17.79	C11H21N3O3	Lys-Pro	1.00	0.00
268.15	8.13	C12H20N4O3	His-Leu	1.00	0.00
219.09	15.53	C7H13N3O5	Asn-Ser	1.00	0.00
294.16	4.88	C15H22N2O4	Leu-Tyr	1.00	0.00
344.24	12.78	C16H32N4O4	Lys-Val-Val	1.00	0.00
358.11	15.64	C12H22O12	melibionate	1.00	0.00
341.23	4.67	C17H31N3O4	Leu-Leu-Pro	1.00	0.00
232.11	15.49	C9H16N2O5	N2-Succinyl-L-ornithine	1.00	0.00
276.06	15.96	C14H12O6	heptaketide pyrone intermediate	1.00	0.00
242.10	14.23	C9H14N4O4	Ser-His	1.00	0.00
269.11	14.79	C10H15N5O4	Asn-His	1.00	0.00
204.07	17.42	C7H12N2O5	Glu-Gly	1.00	0.00
373.21	17.46	C14H27N7O5	Ala-Gln-Arg	1.00	0.00
283.16	17.83	C12H21N5O3	Lys-His	1.00	0.00
388.24	13.85	C16H32N6O5	Leu-Thr-Arg	1.00	0.00
251.08	17.42	C12H13NO5	N-benzoyl-L-glutamate	1.00	0.00
417.26	14.40	C18H35N5O6	Ala-Leu-Lys-Ser	1.00	0.00
192.07	15.21	C6H12N2O5	Ser-Ser	1.00	0.00
260.14	13.26	C11H20N2O5	Glu-Leu	1.00	0.00
252.11	7.32	C12H16N2O4	Ala-Tyr	1.00	0.00
491.20	11.99	C22H29N5O8	Ala-Thr-Trp-Asp	1.00	0.00
321.18	14.48	C15H23N5O3	Phe-Arg	1.00	0.00

204.11	10.39	C8H16N2O4	N6-Acetyl-N6-hydroxy-L-lysine	1.00	0.00
477.26	15.29	C20H39N5O6S	Ile-Lys-Met-Ser	1.00	0.00
422.08	19.12	C12H23O14P	alpha,alpha'-Trehalose 6-phosphate	1.00	0.00
287.20	14.60	C12H25N5O3	Leu-Arg	1.00	0.00
509.17	16.22	C17H31N7O7S2	Glu-Cys-Cys-Arg	1.00	0.00
430.13	11.80	C15H22N6O7S	Asp-Cys-Gly-His	1.00	0.00
517.21	17.65	C19H31N7O10	Asp-Gln-Gln-Gln	1.00	0.00
281.11	14.87	C25H34N6O7S	Glu-Met-Phe-His	1.00	0.00
309.17	17.35	C15H23N3O4	Lys-Tyr	1.00	0.00
317.16	14.34	C13H23N3O6	Leu-Ala-Asp	1.00	0.00
233.10	14.35	C8H15N3O5	Ala-Gly-Ser	1.00	0.00
316.21	14.40	C14H28N4O4	Leu-Lys-Gly	1.00	0.00
337.17	17.23	C15H23N5O4	L-Tyrosyl-L-arginine	1.00	0.00
362.19	17.39	C13H26N6O6	Thr-Ser-Arg	1.00	0.00
256.12	12.95	C10H16N4O4	Thr-His	1.00	0.00
245.17	15.96	C11H23N3O3	Lys-Val	1.00	0.00
293.17	14.47	C15H23N3O3	Lys-Phe	1.00	0.00
206.09	14.25	C7H14N2O5	Thr-Ser	1.00	0.00
374.19	17.64	C14H26N6O6	Lys-Asn-Asn	1.00	0.00
248.10	17.04	C9H16N2O6	Glu-Thr	1.00	0.00
632.27	26.82	C31H36N8O7	Gln-Trp-Tyr-His	1.00	0.00
233.14	18.38	C9H19N3O4	Lys-Ser	1.00	0.00

277.15	15.56	C11H23N3O3S	Lys-Met	1.00	0.00
234.09	17.43	C8H14N2O6	Glu-Ser	1.00	0.00
176.08	14.09	C6H12N2O4	Ala-Ser	1.00	0.00
397.20	13.17	C17H27N5O6	Ala-Asn-Pro-Pro	1.00	0.00
220.11	12.72	C8H16N2O5	N-Acetyl-beta-D-glucosaminyamine	1.00	0.00
304.17	17.73	C12H24N4O5	Lys-Ala-Ser	1.00	0.00
387.25	14.35	C17H33N5O5	Leu-Lys-Gln	1.00	0.00
266.13	7.32	C13H18N2O4	Phe-Thr	1.00	0.00
331.17	11.90	C14H25N3O6	Val-Val-Asp	1.00	0.00
288.15	18.35	C10H20N6O4	Asn-Arg	1.00	0.00
275.16	17.97	C10H21N5O4	Thr-Arg	1.00	0.00
305.15	15.38	C11H23N5O3S	Met-Arg	1.00	0.00
228.11	14.83	C10H16N2O4	(S)-ATPA	1.00	0.00
302.20	15.96	C13H26N4O4	Lys-Val-Gly	1.00	0.00
247.15	17.90	C10H21N3O4	Lys-Thr	1.00	0.00
261.14	18.35	C9H19N5O4	Ser-Arg	1.00	0.00

**Appendix X: Purine metabolism metabolites identified in *Drosophila melanogaster* whole larva and their mean signal intensity**

Mass	RT	Formula	Putative metabolite	17767	18814	18554	CS
175.05	7.96	C <sub>4</sub> H <sub>8</sub> N <sub>4</sub> O <sub>4</sub>	Allantoate	57728.96346	107090.6955	69938.86678	83566.06908
157.04	18.67	C <sub>4</sub> H <sub>6</sub> N <sub>4</sub> O <sub>3</sub>	Allantoin	1729723.557	1352538.551	515909.3323	1168146.404
167.02	13.14	C <sub>5</sub> H <sub>4</sub> N <sub>4</sub> O <sub>3</sub>	Urate	26209132.2	76673121.59	2556547.244	18039415.34
151.03	12.05	C <sub>5</sub> H <sub>4</sub> N <sub>4</sub> O <sub>2</sub>	Xanthine	1.38E+07	17395936.17	11393662.19	18735495.37
135.03	9.20	C <sub>5</sub> H <sub>4</sub> N <sub>4</sub> O	Hypoxanthine	1.50E+07	18425957.35	20254297.4	17465088.36
134.05	9.57	C <sub>5</sub> H <sub>5</sub> N <sub>5</sub>	Adenine	2689326.454	3398902.548	2752861.444	1612219.498
266.09	11.61	C <sub>10</sub> H <sub>13</sub> N <sub>5</sub> O <sub>4</sub>	Adenosine	18228.54369	29087.99755	27229.27468	93743.668
267.07	11.47	C <sub>10</sub> H <sub>12</sub> N <sub>4</sub> O <sub>5</sub>	Inosine	1.72E+08	182474592.2	201427655.8	172103264.9
284.10	13.21	C <sub>10</sub> H <sub>13</sub> N <sub>5</sub> O <sub>5</sub>	Guanosine	1.23E+07	16242326.81	24302747.72	19443905.35

**Appendix XI: Purine metabolism metabolites identified in *Drosophila melanogaster* larval posterior Malpighian tubules and their mean signal intensity**

Mass	RT	Formula	Putative metabolite	17767 PT	18814 PT	18554 PT	CS PT
175.05	7.96	C <sub>4</sub> H <sub>8</sub> N <sub>4</sub> O <sub>4</sub>	Allantoate	35521.33	33719.91	35684.46	979334.40
157.04	18.67	C <sub>4</sub> H <sub>6</sub> N <sub>4</sub> O <sub>3</sub>	Allantoin	711346.76	698017.37	693971.58	694155.39
167.02	13.14	C <sub>5</sub> H <sub>4</sub> N <sub>4</sub> O <sub>3</sub>	Urate	389699.12	44200471.34	8933683.02	2641780.61
151.03	12.05	C <sub>5</sub> H <sub>4</sub> N <sub>4</sub> O <sub>2</sub>	Xanthine	0.00	0.00	0.00	210429.52
135.03	9.20	C <sub>5</sub> H <sub>4</sub> N <sub>4</sub> O	Hypoxanthine	3508715.09	3055189.67	2525559.31	2620326.12
134.05	9.57	C <sub>5</sub> H <sub>5</sub> N <sub>5</sub>	Adenine	1351026.33	1669117.89	1255291.63	4392923.97
266.09	11.61	C <sub>10</sub> H <sub>13</sub> N <sub>5</sub> O <sub>4</sub>	Adenosine	6915.15	3628.62	2992.41	354016.45
267.07	11.47	C <sub>10</sub> H <sub>12</sub> N <sub>4</sub> O <sub>5</sub>	Inosine	3955298.53	4170512.30	2463237.70	3159900.71
284.10	13.21	C <sub>10</sub> H <sub>13</sub> N <sub>5</sub> O <sub>5</sub>	Guanosine	533790.66	564416.28	269604.53	2222183.91

**Appendix XII: Purine metabolism metabolites identified in *Drosophila melanogaster* larval anterior Malpighian tubules and their mean signal intensity**

Mass	RT	Formula	Putative metabolite	17767 AT	18814 AT	18554 AT	CS AT
175.05	7.96	C <sub>4</sub> H <sub>8</sub> N <sub>4</sub> O <sub>4</sub>	Allantoate	3039061.87	32042.76	40877.77	174624.09
157.04	18.67	C <sub>4</sub> H <sub>6</sub> N <sub>4</sub> O <sub>3</sub>	Allantoin	3339911.10	704032.95	700176.25	712290.71
167.02	13.14	C <sub>5</sub> H <sub>4</sub> N <sub>4</sub> O <sub>3</sub>	Urate	3630984.21	1729150.24	0.00	163662.98
151.03	12.05	C <sub>5</sub> H <sub>4</sub> N <sub>4</sub> O <sub>2</sub>	Xanthine	4084857.24	0.00	0.00	312414.71
135.03	9.20	C <sub>5</sub> H <sub>4</sub> N <sub>4</sub> O	Hypoxanthine	4668408.27	3051599.32	2907790.60	2001578.03
134.05	9.57	C <sub>5</sub> H <sub>5</sub> N <sub>5</sub>	Adenine	4841139.22	8025804.88	6039759.09	9533915.79
266.09	11.61	C <sub>10</sub> H <sub>13</sub> N <sub>5</sub> O <sub>4</sub>	Adenosine	799131.12	3795.53	4688.88	191552.99
267.07	11.47	C <sub>10</sub> H <sub>12</sub> N <sub>4</sub> O <sub>5</sub>	Inosine	1192186.75	1947636.92	1862475.11	3321896.20
284.10	13.21	C <sub>10</sub> H <sub>13</sub> N <sub>5</sub> O <sub>5</sub>	Guanosine	563465.89	525769.80	341711.87	2018026.42